Characterization of HIV-1 subtype B near full-length genome sequences identified at the start of the HIV epidemic in South Africa

Adetayo Emmanuel Adegbenga Obasa

Thesis presented in fulfilment of the requirements for the degree of Masters of Science (Medical Virology) at the Faculty of Medicine and Health science, Stellenbosch University

Supervisor

Dr. Graeme Brendon Jacobs

Co-supervisor

Professor Susan Engelbrecht

March 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own original work, that I am the sole owner thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that this work has not been submitted before in its entirety or in part for any degree or examination at this or any other University.

Signature Adetayo Emmanuel Adegbenga Obasa

Date…….

Summary

Summary

South Africa is home to approximately 20.0% of the global Human Immunodeficiency Virus (HIV) infected population. The first reported cases of HIV-1 in the country were described in 1982 amongst the homosexual male population. This was attributed to HIV-1 subtypes B (HIV-1B) and D (HIV-1D). Since the late 1980s HIV-1 subtype C (HIV-1C), spread mainly through heterosexual contact, has been the driving force of the epidemic. To date, only six HIV-1B near full-length genome (NFLG) sequences from South Africa are available in the Los Alamos National Laboratory database (LANL). During this study we retrieved five HIV-1B positive samples from homosexual and bi-sexual males, stored for up to 30 years, from the early 1980s, for further characterization. The NFLG amplification reactions were performed using a modern Polymerase Chain Reaction (PCR) protocol designed to target two overlapping proviral DNA HIV genome fragments, 5.5 kb and 3.7 kb in size, respectively. All positive PCR products were sequenced to characterize the viruses. The sequences were checked and edited manually using Sequencher V5. Multiple sequence alignments were created using Clustal W and Maft V7. The sequences were subtyped using the REGA V3.0, RIP V3.0 and jumping profile Hidden Markov Model (jpHMM) online subtyping programmes. Maximum likelihood phylogenetic trees were drawn using MEGA V6. Four of the five HIV-1 patient sequences were subtyped as pure HIV-1B. One sequence, ZA|85|R605, was characterized as a novel HIV-1 BD recombinant. This is the first NFLG HIV-1 BD recombinant ever described and indicates that recombination events were most likely already happening at the early stage of the South African epidemic. Two patient sequences, ZA|87|R1296 and ZA|87|R459, clusters with HIV-1B sequences from the United States of America (USA). The sequence from patient ZA|87|R68 clusters with a HIV-1B sequence from France and the sequence of ZA|87|R526 clusters with another South African HIV-1B sequence. Homosexual flight stewards, international tourists and migrants from the European and North American countries were most likely responsible for the introduction of the HIV-1B epidemic into South Africa. The findings of this study provides valuable insights from the beginning of the HIV-1 epidemic in South Africa. We highlight the importance of characterizing complete viral genomes from early archival specimens to give a more detailed picture of landmarks of the HIV/AIDS pandemic. We show that NFLG sequencing is an important tool for the identification of recombinant viral strains. This study can form the basis for continued research in our attempt to reconstruct the epidemiology and evolutionary history of HIV in South Africa. The HIV-1 epidemic is dynamic in nature and is constantly changing.

Opsomming

Suid-Afrika is die tuiste van ongeveer 20,0% van die wêreld se menslike immuniteitsgebreksvirus (MIV)-besmette bevolking. Die eerste gerapporteerde gevalle van MIV-1 in die land is beskryf in 1982 onder homoseksuele mans. Dit was toegeskryf aan MIV-1 subtipes B (MIV-1B) en D (MIV-1D). Sedert die laat 1980's oorheers MIV-1 subtipe C (HIV-1 C), hoofsaaklik versprei deur heteroseksuele kontak. Tot hete is slegs ses MIV-1B naby vollengte genoom (NFLG) rye van Suid-Afrika beskikbaar in die LANL MIV databasis. Tydens hierdie studie het ons vyf MIV-1B positiewe monsters van homoseksuele en biseksuele mans, wat gestoor was vir tot 30 jaar, vanaf die vroeë 1980's, verder gekarakteriseer. Die NFLG amplifisering reaksies is uitgevoer met behulp van 'n moderne polimerase-kettingreaksie (PKR) protokol ontwerp om twee oorvleuelende provirale DNS MIV genoom fragmente, 5.5 kb en 3.7 kb in grootte, te teiken. Alle positiewe PKR produkte se DNS volgordes is bepaal om die virusse volledig te karakteriseer. Die volgorde is nagegaan en met die hand geredigeer deur gebruik te maak Sequencher V5. Veelvuldige volgorde roetes is geskep met behulp van CLUSTAL W en Maft V7. Die MIV-1 subtipes is met die aanlyn programme REGA V5, RIP V3.0 en jpHMM bepaal. Die subtipes is bevestig duer maksimum waarskynlikheid filogenetiese bome te trek met behulp van MEGA V6. Vier van die vyf MIV-1 pasiënt volgordes is as suiwer MIV-1B gekarakteriseer. Een volgorde, ZA|85|R605, is gekenmerk as ’n unieke MIV-1 BD rekombinant. Tot ons kennis is dit die eerste en enigste NFLG MIV-1 BD rekombinant ooit beskryf en dui aan dat rekombinasie gebeure reeds in die vroeë jare van die Suid-Afrikaanse epidemie plaasgevind het. Twee pasiënt volgordes, ZA|87|R1296 en ZA|87|R459, is na verwant aan MIV-1B volgordes van die Verenigde State van Amerika (VSA). Die volgorde van pasiënt ZA|87|R68 is na verwant met 'n MIV-1B volgorde van Frankryk en die volgorde van ZA|87|R526 is na verwant aan 'n ander Suid-Afrikaanse MIV-1B volgorde. Homoseksuele vlug bedienaars, internasionale toeriste en immigrante uit die Europese en Noord-Amerikaanse lande was waarskynlik verantwoordelik vir die bekendstelling van die MIV-1B epidemie in Suid-Afrika. Die bevindinge van die studie gee insigte van die begin van die MIV-1-epidemie in Suid-Afrika. Ons beklemtoon die belangrikheid om argiewe MIV-stamme met NFLG volgorde bepalings te karakteriseer aangesien dit vir ons ‘’n meer volledige prentjie van die pandemie kan skep, veral uit die begin jare. Ons wys dat NFLG volgorde bepaling 'n belangrike instrument is vir die identifisering van rekombinante MIV-stamme. Hierdie studie kan die grondslag lê vir voortgesette navorsing om die epidemiologie en evolusionêre geskiedenis van MIV in Suid-Afrika te rekonstrueer.

Dedication

"My help comes from the LORD, the maker of heaven and earth" **(Ps 121: 02).**

I dedicate the thesis to my parents, siblings and Deborah who have always supported and encouraged me. I love you all.

To God be the Glory!!!

Acknowledgements

Firstly, I would like to thank Dr. Graeme Brendon Jacobs who gave me the opportunity to perform this project and kindly received me into his laboratory. Thank you for your unwavering support, financial help and scientific advice. You went above the call of duty just to make a success and get the best out of me. I am so grateful for your continued encouragement and unfailing effort. You will forever remain an inspiration to me. All my gratitude to my co-supervisor Prof. Susan Engelbrecht who supported me throughout the project data analyses and thesis write up. Thank you for guidance and insight into the work reported in this thesis.

Many thanks to all members of the Division of Medical Virology, past and present, for your encouragement, great attitude and help during the course of my study. Many thanks to Mr. Sello Mikasi, Miss Shaheida Isaacs, Miss Olivette Varathan Miss Cynthia Tamandjou, Dr. Suliman Tasnims, Mr. Josiah Gichana and Mrs. Danelle Van Jaarsveldt for taking out time to proof read my thesis. I will also like to appreciate The Redeemed Christian Church of God Christian community for their prayers and moral support. A special thanks to my sweetheart Miss. Deborah Farinloye, thank you for your wonderful support and for always encouraging me to believe in myself.

My sincere appreciation goes all the following my funding agencies: Medical Research Council (MRC), National Research Council (NRF) Poliomyelitis Research Foundation (PRF), Post graduate Merit Bursary (Stellenbosch University) and the National Health Laboratory Services (NHLS). Without your financial support, my project would not have been a success.

Finally, I would also like to appreciate The Bill and Melinda Gates Foundation for Awards that gave me the opportunity to attend the Keystone Symposium international conferences on New Approaches to Vaccines for Human & Veterinary Tropical Diseases Southern Sun Cape Town.

"I have set the Lord always before me:

Because he is at my right hand,

I shall not be shaken." Psalm 16:8

List of Scientific conference output and research visits

Conference Poster presentation

- Virology Africa conference, Cape Town. 30th of November - 3rd December 2015. Poster presentation: "Development of a protocol to analyze near full-length genomes from South Africa".
- 60th annual academic day Stellenbosch University Faculty of Medicine and Health Sciences (Tygerberg campus). 11th of August 2016. Poster presentation: "Near full-length characterization of HIV-1 subtype B identified in South Africa".
- Frontiers of Retrovirology in Erlangen Germany. September 12 – 14 2016. Poster presentation: "Near full-length characterization of HIV-1 subtype B identified in South Africa".

Conference oral presentation

- Annual Pathology Research Day, Stellenbosch University Faculty of Medicine and Health Sciences (Tygerberg campus). 9th of June 2016. "Near full-length characterization of HIV-1 subtype B identified in South Africa".

Conference in attendance

- Keystone Symposia New Approaches to Vaccines for Human & Veterinary Tropical Diseases Southern Sun Cape Town, South Africa. 22nd of May – 26th of May 2016.

List of Research visits

Due to the significance of the study, I was invited by the following collaborators to learn additional techniques:

- Research internship visit to the Institute of Virology and Immunobiology at the University of Wuerzburg, Germany, September to October 2016.
- Research visit to the the Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Sweden. 18th of September to 23rd of September 2016.

LIST OF ABBREVIATIONS

©	Copyright
®	Registered
°C	Degree Celsius
µl	Microliter
A,G,C,T	Adenine, Guanosine, Cytosine, Thymine
AIDS	Acquired Immunodeficiency Syndrome
BIC	Bayesian Information Criterion
BLAST	Basic Local alignment Search Tool
Cpx	Complex
CRF	Circulating Recombinant Form
CTL	Cytotoxicity T Lymphocytes
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of Congo
EDTA	ethylene diamine tetra-acetic acid
Env	Envelope gene
Gag	Group antigen gene
GI	Gamma with Invariant sites
Gp	Glycoproteins
GTR	General Time Reverse
HIV-1	Human Immunodeficiency Virus type 1
HIV-1B	Human Immunodeficiency Virus 1 subtype B
HIV-1C	Human Immunodeficiency Virus 1 subtype C
HIV-1D	Human Immunodeficiency Virus 1 subtype D

HIV-2	Human Immunodeficiency Virus type 2
HMW	High molecular weight
HTLV-III	Human T-cell Lymphotropic Type III
IN	Integrase
jpHMM	Jumping profile Hidden Markov Model
Kb	Kilo base pairs
KS	Kaposi `s sarcoma
KZN	Kwa Zulu Natal
LANL	Los Alamos National Laboratory
LAV	Lymphadenopathy Associated Virus
LMICs	Low Middle Income Class settings
LTR	Long Terminal Repeats
MAFT	Mulitple Alignment using Fat Fourier
MEGA	Molecular Evolutionary Genetics Analysis
ML	Maximum Likelihood
mM	Millimolar
MMWR	Morbidity and Mortality Weekly Report
MRC	Medical Research Council
MSM	Men who have sex with men
Nef	Negative Factor
NFLG	Near Full Length Genome
NFLG	Near Full Length Genome
ng	Nanogram
NHLS	National Health Laboratory Services

NJ	Neighbour-Joining
NRF	National Research Foundation
NGS	Next Generation Sequencing
NTC	No Template Control
ORF	Open Reading Frames
PBS	the primer binding site
PBMC	Peripheral blood mononuclear cells
PCP	*Pneumocystis carinii* Pneumonia
PCR	Polymerase Chain Reaction
Pol	Polymerase gene
PR	Protease
PRF	Poliomyelitis Research Foundation
Rev	Regulator of Virion Expression
RIP	Recombination Identification Program
RNA	Ribonucleic acid
RT	Reverse Transcriptase
RPM	Revolution per minute
SIV	Simian Immunodeficiency Virus
SU	Surface glycoproteins
Tat	Transcriptional transactivator gene
TM	Transmembrane protein
tRNA	Transport ribonucleic acid
Txt	Text
™	Trademark

URF	Unique Recombinant Forms
USA	United States of America
UK	United Kingdom
URAI	Unprotected Receptive Anal Intercourse
Vif	Viral Infectivity Factor gene
Vpr	Viral Protein R gene
Vpu	Viral Protein U gene
WC	Western Cape

Table of Content

List of Figures

List of Tables

CONTENT

CHAPTER 1

Introduction and literature review

INTRODUCTION

1.1.1. Brief introduction

The Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS). There are two phylogenetically distinct types, that are referred to as HIV type 1 (HIV-1) and HIV type 2 (HIV-2) (Hirsch *et al.*, 1989; Robertson *et al.*, 2000). HIV-1 is classified into four main groups namely; Group M (major), Group N (non-M), Group O (outlier) and Group P. HIV-1 group M has nine non-recombinant subtypes designated with letters A-D, F-H, J and K (Robertson *et al.*, 2000; Aldrich and Hemelaar, 2012). HIV-1 subtype B (HIV-1B) is a subtype that belongs to HIV-1 Group M. Scientific reports have also described several circulating recombinant forms (*CRFs*), unique recombinant forms (*URFs*) as well as second generation recombinants (*SGRs*) of group M (Aldrich and Hemelaar, 2012). The geographical distribution of group M recombinants (78 *CRFs* and several *URFs*) throughout the world is not equal (http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html accessed 2016 August 30; Aldrich and Hemelaar, 2012). HIV-2 is restricted to West African countries (Aldrich and Hemelaar, 2012; Essex and Mboup, 2002). *CRFs* are formed from the recombination of two viral subtypes in at least three epidemiologically unlinked individuals and subsequently further transmitted in the population (Tebit and Arts, 2011; Lau and Wong, 2013). As an example, recombination between HIV-1 subtypes B and F is referred to as CRF_12BF; with the number 12 representing the order in which the *CRF* was described. *URFs* are known to emerge from the recombination of two or more subtypes which do not meet the CRF criteria (https://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes-more.html accessed January 10 2017). The long infection period, in conjunction with population demographics, has led to the rapid evolution of HIV-1 group M (Aldrich and Hemelaar, 2012). This has resulted in a complex classification, worldwide spread and strains intermixing (recombinants) (Aldrich and Hemelaar, 2012). From the beginning of the HIV pandemic, an estimated 78 million people have been afflicted with AIDS (UNAIDS, 2015). Approximately 35 million people have died of AIDS-related illnesses. Sub-Saharan Africa is the most affected region and accounts for almost 70% of the total new infections globally (UNAIDS 2015).

South Africa has experienced the most devastating effect of HIV/AIDS (UNAIDS, 2015). Evidence suggest that the current burden caused by HIV in South Africa is predominantly driven by heterosexual transmission of the virus (Fraser-Hurt *et al.*, 2011; Zuma *et al.*, 2016). HIV-1 (group M) subtype C (HIV-1C) predominates in South Africa. The initial epidemic in the country, from the early 1980s, were caused by HIV-1 subtype B (HIV-1B) and D (HIV-1D), spread mainly in the homosexual population (Engelbrecht *et al.*, 1995; Middelkoop *et al.*, 2014). Although in minority, HIV-1B is still circulating in the country and our laboratory have previously produced evidence of HIV-1BC recombinant strains circulating in the Western Cape Province of South Africa (Jacobs *et al.*, 2014). The widely accepted hypothesis is that the worldwide introduction and spread of HIV-1B is because of human migration from Democratic Republic of Congo (DRC) in Africa into the Haitian population. Human movements played a fundamental role in the worldwide dissemination of HIV-1B, which led to the viral spread into the diverse population (Gilbert *et al.*, 2007; Junqueira and de Matos Almeida, 2016). **Figure 1** displays the worldwide spread of HIV-1B.

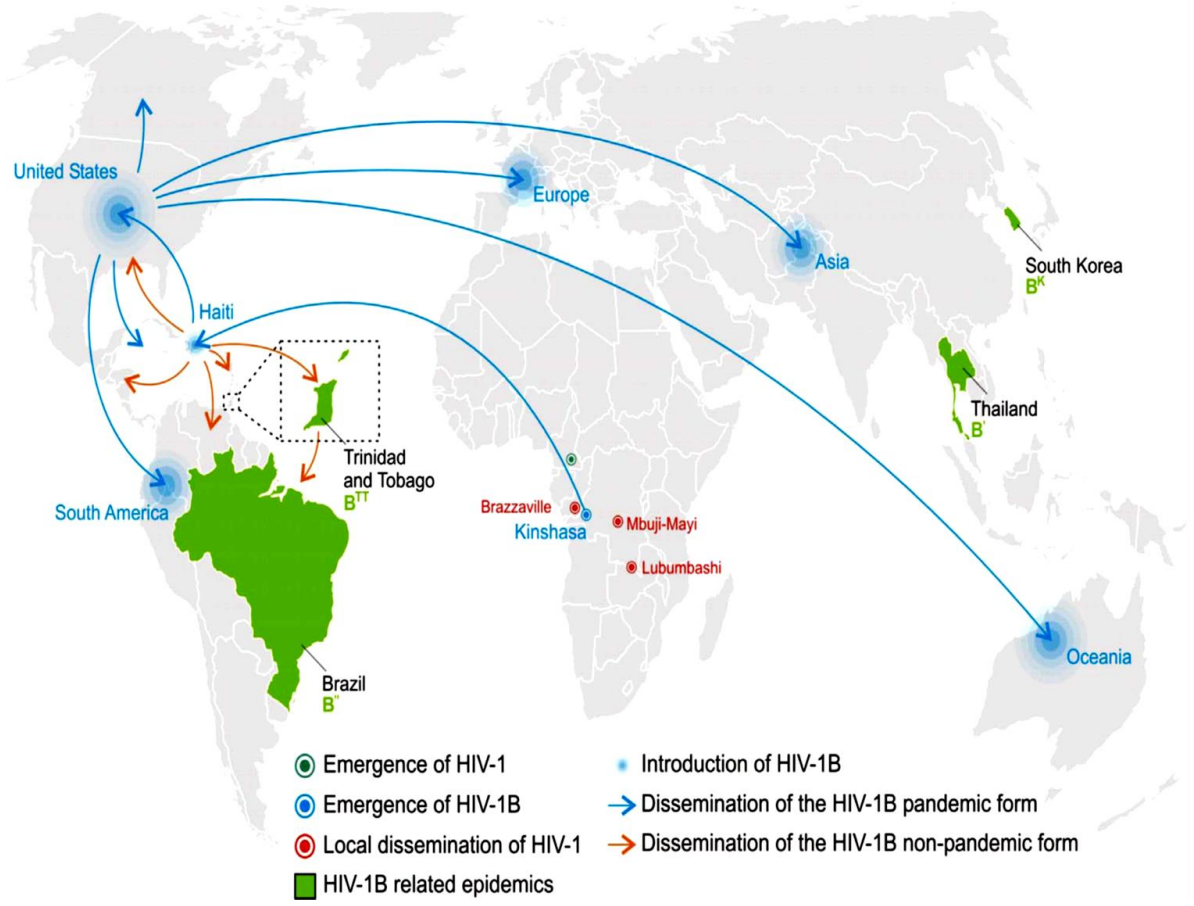


Figure 1: The global dissemination of HIV-1 subtype B (HIV-1B). This diagram shows how HIV-1B originated from the capital city of Democratic Republic of the Congo (DRC) Kinshasa in Africa. The orange lines indicate the direction of the HIV-1 subtype B non-pandemic lineage dissemination. The Green regions indicate countries where specific lineages are circulating (Junqueira and de Matos Almeida, 2016).

1.1.2. HIV-1 spread in South Africa

In South Africa two separate epidemics have been described; as shown in **Figure 2.** The first reported cases of AIDS in South Africa occurred in 1982, which was initially spread by homosexual men (Kuster *et al*., 1994; Sher, 1989; Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997; Loxton *et al.*, 2005; Jacobs *et al.*, 2007). Until 1987, HIV-1 was still diagnosed almost exclusively in men. This was later identified as HIV-1B and D (Engelbrecht *et al.*, 1995; Loxton *et al.*, 2005; Jacobs *et al.*, 2007).

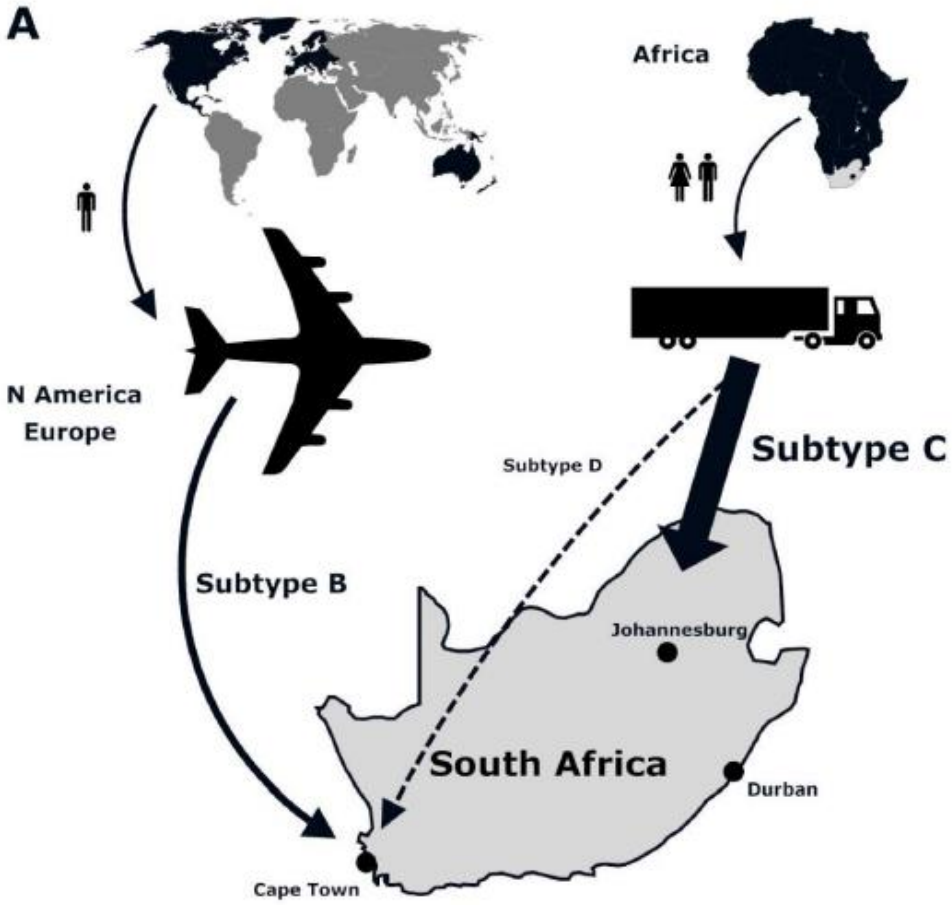


Figure 2: The Spread of HIV-1 in South Africa. This diagram illustrates the critical role of human migration played in the spread of HIV-1 in South Africa (adopted from Prof. Susan Engelbrecht).

The second stage of the epidemic was first described in the year 1988 when an increasing number of HIV-1C viral sequences were diagnosed in heterosexual females. In 1992, the number of new cases in females was roughly equivalent to those in males (Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997). HIV-1C has been responsible for the majority of infections in South Africa (Jacobs *et al.*, 2014).

1.2. AIM OF THE STUDY

The current study is aimed to fully characterize HIV-1B near full-length genome (NFLG) virus sequences from the early South African epidemic from the 1980s. Until date, we only have six HIV-1B NFLG sequences from South Africa and one other African NFLG sequence from Gabon in the HIV database Los Alamos National Laboratory (LANL). With advanced molecular techniques, we now have the opportunity to expand our knowledge / understanding of the molecular characteristics of the start of the HIV epidemic in South Africa. Therefore, NFLG sequences generated from this work will serve as an update for the HIV database. NFLG characterization of HIV-1B sequences from the beginning of the epidemic, with the previously identified subtype B, may shed light on how HIV evolution in South Africa.

.

1.3. LITERATURE REVIEW

In the following section, retroviruses, the history of HIV-1, origin of the pandemic, morphology, organisation of the genome, life cycle as well as subtype distribution of HIV-1B worldwide and in South Africa will be briefly discussed. In addition, a brief introduction of the phylogenetic analyses that are relevant to the project will be reviewed.

1.3.1 Retroviruses

Retroviruses received their name in 1974 (Baltimore *et al*., 1974; Weiss, 2006). A retrovirus possesses a unique enzyme called reverse transcriptase (RT). This unique enzyme uses viral Ribonucleic acid (RNA) as a template to make a Deoxyribonucleic acid (DNA) copy, which is then integrated into its host genomic DNA (Barre-Sinoussi *et al*., 1983). Retroviruses are also diploid, which means they contain two nuclei acid copies, it possesses a high recombination rate when different parental genomes are in the same virus particle (Robertson *et al.,* 1995). Retroviruses were initially of little concern as they were not associated with humans and were only identified in primates. The perceived opinion changed with the discovery of human-associated retroviruses. The classification of human retroviruses is depicted in **Figure 3.**

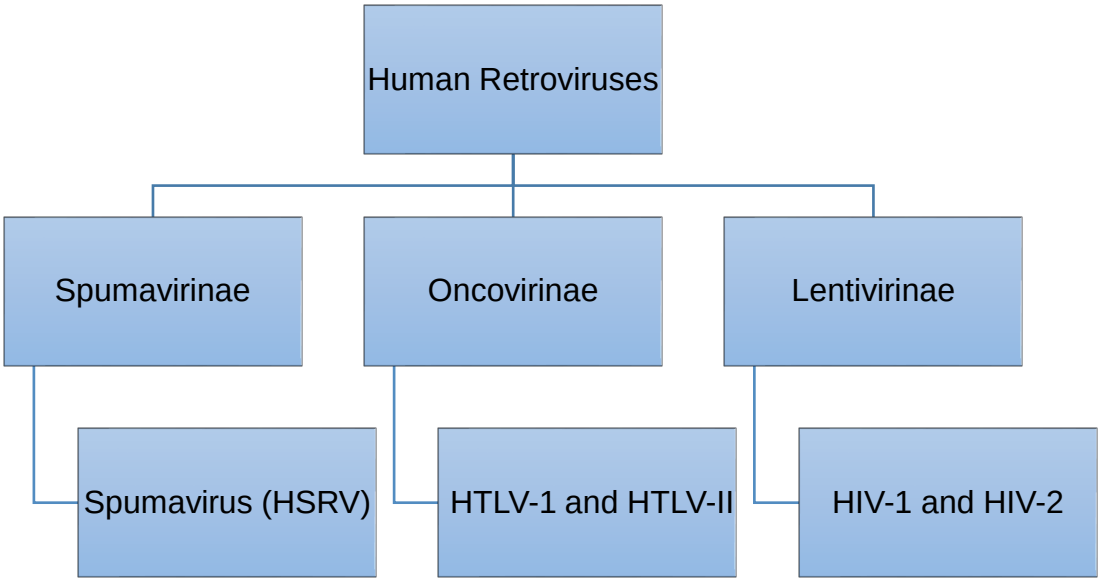


Figure 3: Human Retrovirus classification. Retroviruses are classified into Spumavirinae, Onconovirinae and Lentivirinae. HIV-1 and HIV-2 belong to the genus Lentivirinae (Robertson *et al*., 1995).

Retroviruses can be divided into three families: Lentivirinae (HIV-1 and HIV-2), Oncovirinae (HTLV-I and HTLV-II) and Spumavirinae. The first human retrovirus isolate was the human spumavirus (HSRV) (Barre-Sinoussi *et al.*, 1983). The human spumavirus can be isolated from a wide range of mammal's species and are found worldwide (Bobkov *et al.*, 1998; Ghersi *et al*., 2015). The Oncovirinae family consists of Human T-cell leukemia virus Type I or Type II (HTLV-1 and HTLV-II) and has the ability to induce cancerous conditions such as carcinoma (Ammann *et al.*, 1983; Fahey *et al.*, 1984). HTLV-I are thought to have shared the same origin and transmission as HIV-1. HTLV-II was first isolated from a patient's cells who presented with clinical symptoms of a rare form of leukemia (Gallo, 1984). HTVL-II was later hypothesized to have originated from Africa and subsequently spread to other parts of the world. It is mostly associated with intravenous drug users (IDUs) (Vandamme *et al*., 1998). The Lentivirinae family of retroviruses includes HIV-1 and HIV-2. In primates, they are termed Simian Immunodeficiency Virus (SIV). HIV-2 originated from Sooty mangabey and is found exclusively in West Africa (Clavel *et al*., 1987). **Figure 4** indicates the migration of SIV into humans and its diversity in both primates and homosapiens.

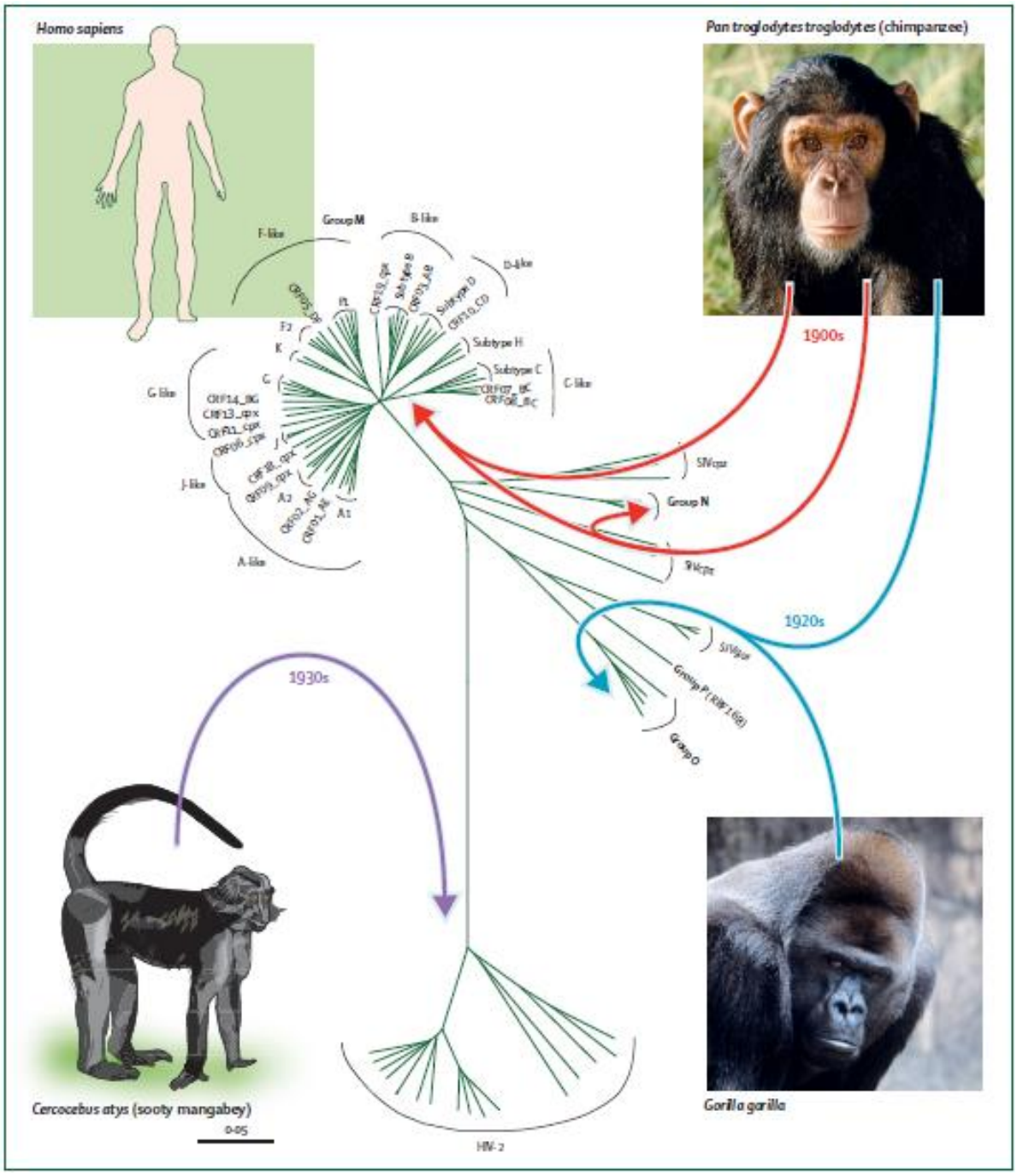


Figure 4: Phylogenetic tree illustrating the relationship of HIV-1 and HIV-2 in humans with respect to primate SIV. The diagram shows the pattern of cross-species transmission and HIV-1 genetic diversity classified into groups M, N, O and P. HIV-2 and SIVs. This Figure was adoption from (Tebit and Arts, 2011).

1.3.2 The history of HIV-1

AIDS was first described in the early 1980's in the United States of America (USA) amongst homosexual males. Patients presented with a rare opportunistic microbial infection caused by pneumocystis *carinii* pneumonia and some later develop Kaposi's sarcoma (MMWR 30, 1981; MMWR 31, 1981; Gottlieb *et al.*, 1981; Friedman-Kien, 1981; Masur *et al.*, 1982). Human Immunodeficiency Virus type 1 (HIV-1) was subsequently designated as the etiological agent of AIDS by independent research groups in France and the USA in 1983 and 1984, respectively (Barre-Sinoussi *et al.*, 1983; Popovic *et al.*, 1984).

The earliest clinical evidence of HIV-1 infection was from samples collected in 1959 from a Kinshasa male and also from Norwegian sailors dated in 1971 and 1976 (Zhu *et al.*, 1998; Froland *et al.*, 1988). A second diagnosis of HIV-1 was confirmed in Kinshasa collected during 1960, showed variation in the HIV-1 *env* gene compared to the 1959 strain and suggested that HIV had already been diversifying and underwent molecular evolution in humans before 1960 (Korber *et al.*, 2000; Worobey *et al.*, 2008). Based on this data, the latest introduction of HIV into humans can be dated to around 1900 (1902 to 1921), 20 years earlier than previously estimated (Korber *et al.*, 2000; Worobey *et al.*, 2008). The origin of HIV-1 has therefore been a topic of extensive debate in the past (Gao *et al.*, 1999; Worobey, 2004). HIV-1 is closely related to SIV from chimpanzees, mainly to strains from the subspecies, *Pan troglodytes troglodytes* (Gao *et al.*, 1999).

1.3.3 The origin of the HIV pandemic

HIV originated from multiple zoonotic transmissions of SIV from non-human primates into humans in West Central Africa, occurring over decades which then resulted in the divergence of the virus (Tebit and Arts, 2011; J., 2012). The transmission most likely occurred through hunting and slaughtering of primates for bush meat and the capturing of simians as pets (Hahn *et al.*, 2000). Phylogenetic evidence has revealed that at least three zoonotic transmissions occurred resulting in the HIV-1 M, N and O groups (Figure 1.1) (Sharp *et al.*, 2001; Lemey *et al.*, 2004; Keele *et al.*, 2006). HIV-1 group M is responsible for the majority of the infections worldwide (Lemey *et al.*, 2004; Hemelaar, Gouws, Peter D Ghys, *et al.*, 2011). Researchers have also found evidence of an O-like virus in gorillas (SIVgor), which further supports the theory of a common ancestor being shared between primates and homosapiens (Keele *et al.*, 2006). Furthermore, group O and P viruses share similarities and suggest that the natural hosts of gorilla infections were Pan troglodytes chimpanzees (Plantier *et al.*, 2009). Group O and P is likely to have been derived from independent transmissions

of SIV from gorillas (SIVgor) to humans. This seems likely for group P as it clusters closely with SIVgor in phylogenetic analysis. Thus far only two people have been identified with HIV infections from group P and occurred so far in Yaoundé, Cameroon (Plantier *et al.*, 2009; Vallari *et al.*, 2011). HIV-2 is endemic in West Africa and it has limited spread outside West Africa (Clavel *et al.*, 1987). HIV-2 is characterized by a longer asymptomatic state, lower viral loads and lower mortality rates as compare to HIV-1 (Clavel *et al.*, 1987).

1.3.4 HIV-1 morphology and genome organisation

The HIV-1 virion is cylindrically shaped and surrounded by an external bi-lipid layer membrane, as shown in **Figure 6**. It is approximately 120nm in diameter (Gallo, 1984; Novitsky *et al.*, 1999). The viral Envelop is a bi-lipid layer derived from budding of the virus from its host cell (Gallo, 1984). The Envelope protein is covered with the outer trimeric surface glycoprotein (gp120) and transmembrane glycoprotein (gp41) as a result of the precursor protein (gp160) (Turner and Summers, 1999). A Matrix (MA) protein (p17) appears on the inside of the lipid bi layer shell which forms the inner surface (Turner and Summers, 1999). Capsid (CA) protein (p24) layer forms the conical capsid (Girard *et al.*, 2011). The genome is a full length diploid linear ribonucleic acid (RNA) (Freed, 2015). HIV-1 enzymes inside the genome are Protease (PR), Reverse Transcriptase (RT) and Integrase (IN). **Figure 7** displays the HIV-1 genomic organisation. HIV-1 also has structural proteins which are called; Trans-Activator of Transcription (*Tat*) and Regulatory of virion expression (*Rev*). Regulatory genes are also classified as accessory genes which include Viral protein unique (*vpu*), Viral protein R (*vpr*), Viral infectivity factor (*vif*) and Negative regulatory factor (*nef*) (Girard *et al.*, 2011; Freed, 2015). The gene functions are outlined in **Table A.**

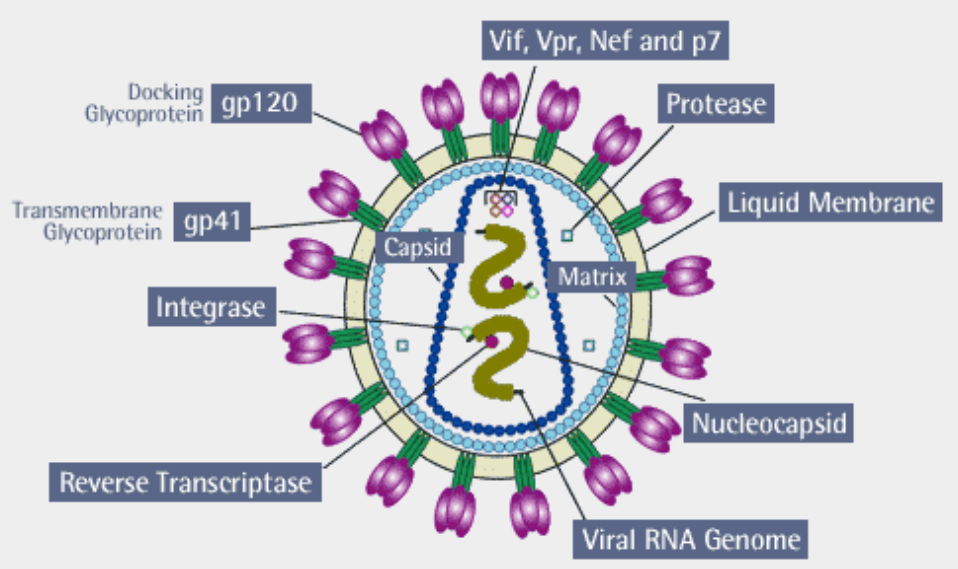


Figure 5 A: Schematic diagram of the HIV-1 virion. The diagram gives a transactional view of the HIV virion and displays the virus matrix encloses the capsid, which protects the two copies of genomic RNA, reverse transcriptase, integrase and protease. The Envelope consists of protruding glycoprotein gp120, which stems from the fusion protein gp41. This figure was obtained from http://www.infohow.org/science/biology-ecology/hiv-viron/, accessed 2016-08-12.

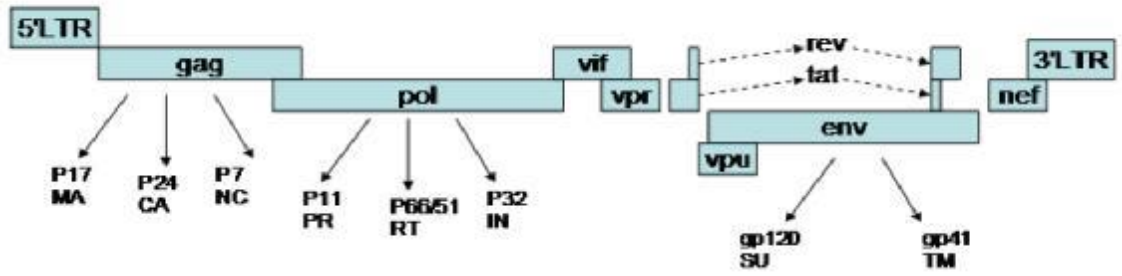


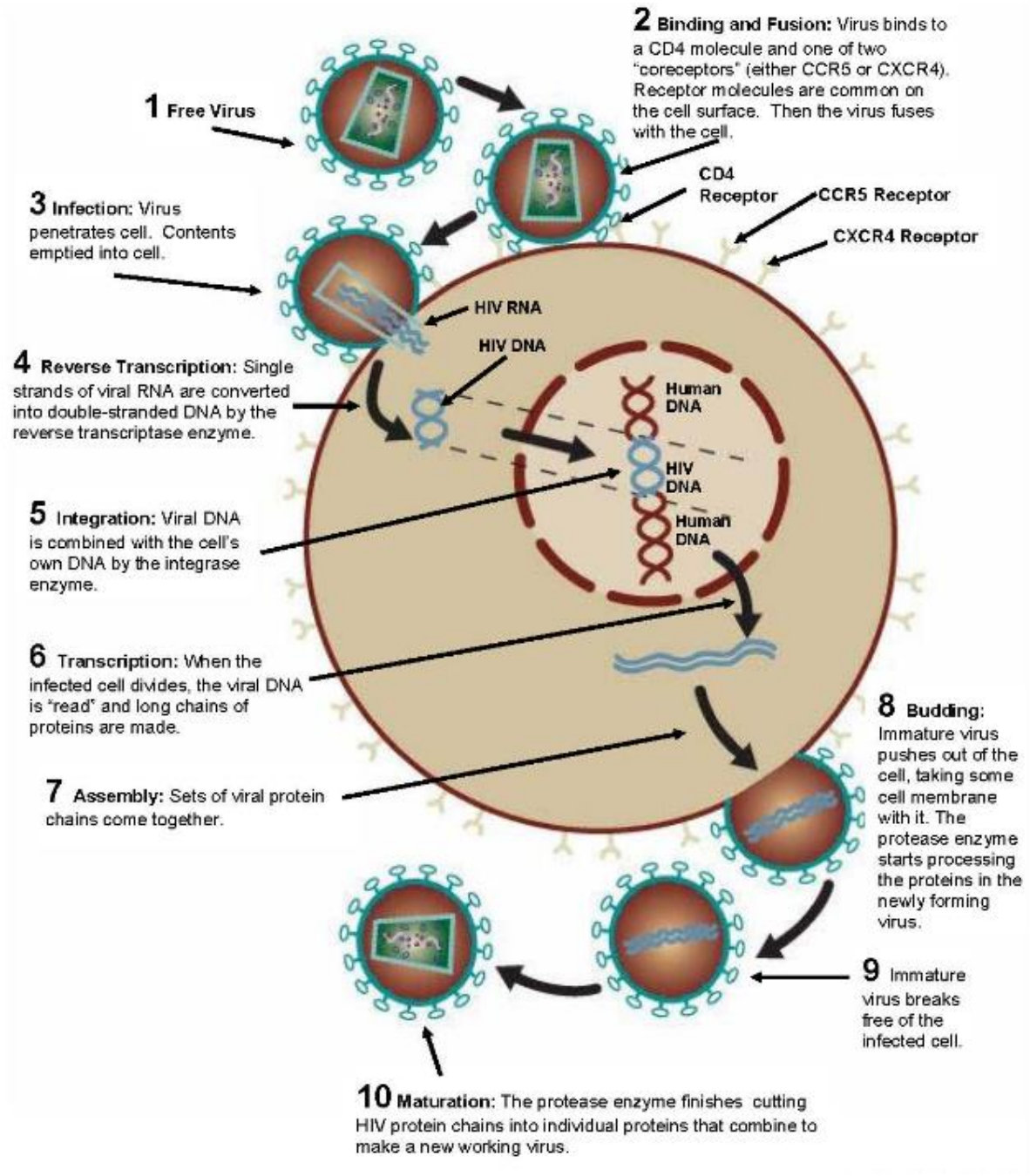
Figure 6: Outlay of HIV genes, and morphology: The image illustrates the genetic composition of HIV and the proteins produced by each gene. MA-Matrix (p17); CA-Capsid (p24); NC-Nucleocapsid (p7); PR (P10); RT-Reverse transcriptase (p66), IN-Integrase (p32), SU- Surface subunits (gp120), TM- transmembrane subunit (gp41) obtained from the (Costin, 2007).

Table A: HIV gene names proteins and their functions. Table A contains information compiled from (Graves *et al.*, 1988; Jaskólski *et al.*, 1991; Stanley *et al.*, 2008; Mascarenhas and Musier-Forsyth, 2009; Kopietz *et al.*, 2012; Solbak *et al.*, 2013; Freed, 2015). The expression of the full genome requires multiple frame shifts to synthesize mRNA from overlapping genes (Holmes, Zhang and Bieniasz, 2015).

Protein	**Protein name**	**Size (kDa)**	**Functions**
Gag (Structural Protein)	Gag	55	Precursor protein for MA, CA and NC
	Matrix (MA)	17	Attaches to the inner cell membrane. Binds RNA to the plasma membrane
	Capsid (CA)	24	Encloses the viral genome
	Nucleocapsid (NC)	6-7	Stabilizes the viral RNA genome
Pol (Structural protein)	Protease (PR)	11	Proteolytic processing of the Gag and Gag-Pol polyprotein into mature chains
	Reverse transcriptase (RT)	66 (p66) 51 (p51)	Conversion of viral RNA into viral DNA
	Integrase (IN)	32	Integrate viral DNA into genomic DNA
Env (Structural protein)	gp160	160	Precursor for gp120 and gp41
	gp120	120	Co-receptor (CCR5/CXCR4)binding
	gp41	41	Help with the fusion of virus membrane with the host cell membrane
Accessory proteins	Negative regulatory factor (Nef)	27-35	Optimizes T-cell stimuli and down regulates CD4 regulation
	Viral protein R (Vpr)	14	Enhance HIV-1 replication
	Viral protein unique (Vpu)	16	Enhances interaction with Tetherin
	Viral infectivity factor (Vif)	23	Inhibits viral activities of host cell enzymes APOBEC
Regulatory proteins	Trans-Activator of Transcription (*Tat*)	14	Initiates transcription by binding integrated viral DNA
	Regulatory of virion expression (*Rev*)	18	Regulates viral replication by a controlled translocation of unspliced and spliced viral RNA from the nucleus into the cytoplasm

1.3.5 The HIV-1 life cycle

Entry of HIV into the host cell depends on the cell surface receptors molecules CD4 and co-receptors CCR4 and CCR5 (Freed, 2015). Once inside the cell, HIV is uncoated and viral RNA is reverse transcribed by the Reverse Transcriptase (RT) enzyme into pro-viral double stranded DNA and subsequently transported to the cell nucleus (Pomerantz and Horn, 2003). RT inhibitors can block this process. At this stage Integrase (IN) aids integration of viral DNA into host genomic DNA and renders the infection irreversible (Freed, 2015). IN inhibitors can block this process and prevent the infection step. Finally, the virus can synthesize the viral protein by its effective use of host machinery and enzymes. New virions are produced by the assembly and budding through the infected cell membrane (Pomerantz and Horn, 2003).

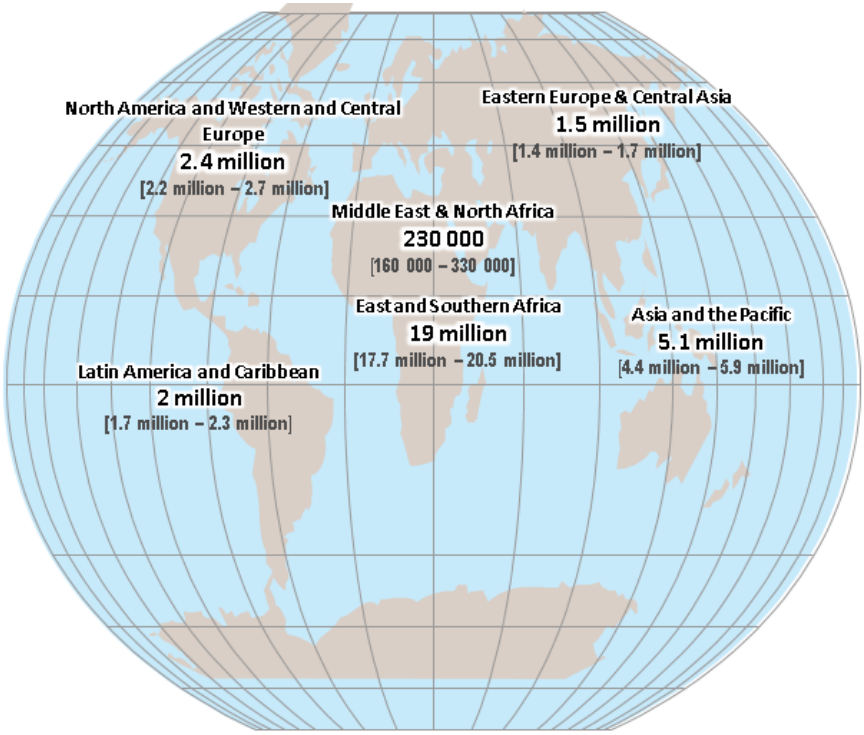


Source: http://www.aidsinfonet.org

Figure 7: The HIV lifecycle. The stages of viral lifecycle shown above. Thin, dark arrows show entry and integration. Curved and slightly bent arrows show early and late replication. 1- A piece of genetic material surrounded by protein. 2- Adsorption to the CD4 receptor (CXCR4 or CCR5). 3- Fusion and un-coating of a viral genomic dimer. 4- Reverse transcription. 5- Integration of pro-viral DNA into host genome using the integrase enzyme. 6- Transcription of viral DNA Nuclear import of PIC. 7- Assembly of viral chain 8- Viral budding, protease enzyme starts the formation of a new virus. 9- Breaking off the immature virus. 10- Viral maturation and formation of the new virus (Pomerantz and Horn, 2003; Freed, 2015).

1.3.6 The epidemiology of the HIV pandemic

The year 2016 marks 35 years of the AIDS pandemic since its discovery in 1981. In 2015, the United Nations AIDS foundation (UNAIDS) estimated that 36.7 (34.0 – 39.8) million people are currently living with HIV, with 2.1 (1.8 – 2.4) million new HIV infections in 2015 globally (UNAIDS, 2015). Due to the massive antiretroviral therapy (ART) upscale, there has been a decline in AIDS-related deaths ranging from 1.2 (990,000 – 1.4) million in 2014 to 1.1 (940,000 – 1.3) million in 2015 (UNAIDS, 2015). The recent decline has been encouraging to the public health society. Southern and East Africa is the worst affected by HIV with approximately 20.5 million people infected with the virus (UNAIDS 2015).



Source: http://www.unaids.org/

Figure 8 : Illustration of the AIDS pandemic in 2015. The map indicates the global spread of HIV and indicates the areas most affected namely East and Southern Africa. Patients include Adult and Children (UNAIDS 2015).

1.3.7 The epidemiology of HIV-1 in South Africa

South Africa currently has the largest AIDS epidemic profile in the world as a result of poverty, diverse cultural beliefs, females subjected to sexual abuse, limited primary health care and education in rural areas (Mayosi *et al.*, 2014; UNAIDS, 2015). By December 2015, approximately 7.0 (6.7 – 7.4) million people were infected with HIV in South Africa. HIV-1 was most prevalent in the KwaZulu-Natal (KZN) province with a prevalence of 27.6%. The Western Cape (WC) province has the overall lowest prevalence of approximately 9.2% three times lower than KZN (Delva and Karim, 2014). In South Africa, only 3.4 million people are currently receiving ART (UNAIDS 2015). Patients receiving ART have a greater improvement in terms of life expectancy and morbidity (UNAIDS 2015).

1.4 Distribution of HIV-1 group M subtypes

HIV-1 group M has a high genetic variability and HIV-1 subtypes are not evenly distributed worldwide (Aldrich and Hemelaar, 2012). Most of the Group M subtypes are genetically linked and human migration plays a vital role in the dissemination and spread between neighbouring countries (Gao *et al*., 1998). Subtype A was first identified in the sub-Saharan Africa (Tebit and Arts, 2011). It has also been identified in Russia and many Western countries (Tebit and Arts, 2011). HIV-1B is more predominant in Europe, North America, Southeast Asia, Middle East and Australia (Junqueira and de Matos Almeida, 2016; Magiorkinis *et al.*, 2016). HIV-1C is most dominant subtype in Southern Africa, similar to East Asia (Jacobs *et al.*, 2014; Wilkinson *et al.*, 2015). HIV-1D was first isolated in a Zairian student in Alabama and the strain was fully sequenced in the USA which highlights the fact that almost all HIV-1 subtypes are of Africa origin (Graves *et al.*, 1994; Puren, A, 2002). HIV-1D has since been isolated in Uganda, Kenya, Tanzania South Africa and Cuba (Thomson, Pérez-álvarez and Nájera, 2002). Subtype F has been described in Belgium, Romania and South America (Thomson, Pérez-álvarez and Nájera, 2002). Subtype J and H have been isolated in Central Africa, Cyprus, Greece and Cuba (Thomson, Pérez-álvarez and Nájera, 2002; Loxton *et al.*, 2005; Tebit and Arts, 2011).

1.4.1 The emergence of the HIV-1B epidemic

The HIV-1B common ancestral strain most likely originated in Kinshasa (Faria *et al.*, 2014). Phylogenetic and historical evidence indicates that HIV-1B has been circulating in the region since 1944 (1935 – 1951) (Faria *et al.*, 2014). The spread of the subtype B lineage and epidemic initially began outside Africa around 1966 (1962 – 1970) (Gilbert *et al.*, 2007). In the 1960s, approximately 4500 Haitian professionals were sent to work in Congo. The majority of the immigrants were teachers and bureaucrats (Junqueira and de Matos Almeida, 2016). The subsequent political crisis and Belgian congo in the city of Kinshasa led to migration of the Haitian professionals back to Haiti (Piot *et al.*, 1984; Pepin, 2011; Crowder, 1984). In 1979, Haitians analyzed hospitalized patients with AIDS-related symptoms; the first AIDS case was only confirmed in 1983 (Malebranche *et al.*, 1983; Pape *et al.*, 1983). Initially, the HIV-1B viral dissemination in Haiti was predominantly identified with the men who have had sex with men (MSM) risk group (Wheeler and Radcliffe, 1994; Gilbert *et al.*, 2007; Worobey *et al.*, 2008; Junqueira and de Matos Almeida, 2016). Over a period of time the nature of the epidemic has changed and reached the heterosexual population and recipients of blood transfusion in Haiti (Junqueira and de Matos Almeida, 2016).

1.4.2 Spread and global distribution of HIV-1B

HIV-1B occupies a vital position in various continent epidemiological profiles and successfully been dispersed across the globe (https://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp accessed on the 25th of August 2016). It is currently the only subtype circulating in several countries of each continent all over the world (https://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp accessed on the 25th of August 2016). Figure 9 shows that the LANL HIV database has over 609128 sequences in total and has approximately 339370 HIV-1B sequences accounts for 55.71% across the globe. Only seven of these full-lengths sequences are from Africa (LANL accessed 2016 August 30). Global distribution of HIV-1B have been reviewed by Lukashov *et al.*, 1996; Bobkov *et al.*, 1998; Robertson *et al.*, 2000; Liebert *et al.*, 2000; Nabatov *et al.*, 2002; Guimara *et al.*, 2007; Arán-Matero *et al.*, 2011; Hemelaar, Gouws, Peter D Ghys, *et al.*, 2011; Junqueira *et al.*, 2011; Pagán and Holguín, 2013; Cabello, Mendoza and Bello, 2014; Junqueira and de Matos Almeida, 2016.

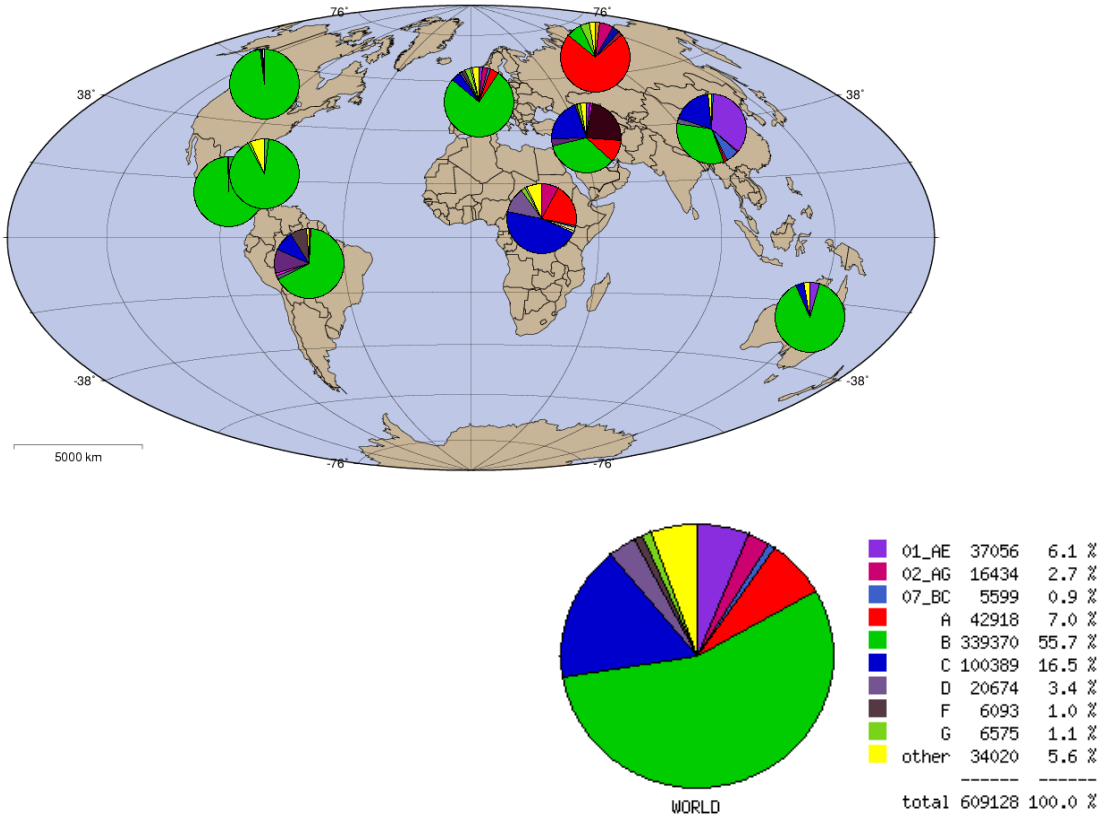


Figure 9: Illustrates the current distribution of HIV-1 subtype sequences as found in the LANL HIV database. Most recent subtype sequences according to distribution, population and spread

across the continent. The green portion represents the worldwide distribution of HIV-1 subtype B across the globe. Adopted from the LANL HIV database (https://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp, accessed 2016 August 30).

1.4.3 HIV-1B pandemic and non-pandemic clades

The earliest evidence of HIV-1 introduction into the USA emerged approximately in 1969 (Junqueira and de Matos Almeida, 2016). Several studies identified a high HIV-1B prevalence in MSM amongst the Americans and Haitians living in America (Laverdiere and Lavallee, 1983; Selik *et al*., 1984; Robbins *et al.*, 2003; Gilbert *et al.*, 2007; Junqueira and de Matos Almeida, 2016). The HIV-1B rate of infection has increased exponentially amongst the MSM population that formed the bridge between both countries (Kuiken *et al.*, 2000; Pape *et al*., 1983). In the mid-1970s, the capital city of Haiti, Port-au-Prince, was a popular tourist destination for homosexual Americans, mostly from the New York Metropolitan area, in search of sex tourism (Pepsin, 2011).

In the United States, the rate of infection was higher amongst homosexuals than heterosexuals (Paraskevis *et al.*, 2009; Beyrer, Baral and Griensven, 2012; Beyrer *et al.*, 2013; Patel *et al.*, 2014). This could be due to the higher incidence of anal intercourse as opposed to vaginal intercourse, as most infections in MSM are transmitted through unprotected receptive anal intercourse (URAI) (Patel *et al.*, 2014). Social connection and the international relation between countries played a vital role in the spread of HIV-1B to most parts of the world, including South Africa. The introduction of the virus because of high-risk behavior explains the successful establishment of HIV-1B epidemic in other parts of the world. (Kuiken *et al.*, 2000; Perrin, Kaiser and Yerly, 2003).

In the early 1970s, a secondary non-pandemic outbreak originated in Hispaniola Island. By the 1980s, it had spread across the majority of the South and Central American countries, including Trinidad and Tobago, Jamaica, Mexico, Venezuela, Panama, Columbia, Ecuador, El Salvador, Honduras Suriname and Brazil (Cabello, Junqueira and Bello, 2015; Junqueira and de Matos Almeida, 2016). These clades reflect virus circulation in culturally related countries, which gives rise to transmission networks. The transmission in most cases often results in dead-end infection due to the combination of chance and multifaceted socio-ecological factors (Cabello, Junqueira and Bello, 2015).

1.4.4 The HIV-1B epidemic in South Africa

The origin of HIV-1 in South Africa began with HIV-1 subtype B and D in the homosexual risk group (Sher, 1989; Becker, De Jager and Becker, 1995; Engelbrecht *et al.*, 1995). HIV-1B was introduced into the country by homosexual males that worked as flight stewards and tourist traveling within North America and European countries (Engelbrecht *et al.*, 1995). The second phase of HIV infection occurred in the late 1980s where the epidemic was diagnosed amongst heterosexuals (Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997; Harmelen *et al.*, 1999). In the early 1980s subtype B is almost exclusively seen in homosexual males in South Africa (Sher, 1989; Becker, De Jager and Becker, 1995; Engelbrecht *et al.*, 1995; Loxton *et al.*, 2005; Jacobs *et al.*, 2007). There have been cases of heterosexual HIV-1 B described within our laboratory (Dr. Jacobs, personal communication).

Currently, HIV-1C is the dominant subtype circulating in South Africa. HIV-1 has been studied by Becker, De Jager and Becker, 1995; Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997; Harmelen *et al.*, 1999, 2001; zur Megede *et al.*, 2002; Treurnicht *et al.*, 2002; Gordon *et al.*, 2003; Hunt *et al.*, 2003; Rousseau *et al.*, 2006; Jacobs *et al.*, 2008, 2014; Wilkinson and Engelbrecht, 2009.

Almost four decades since the first recognition of HIV-1 in South Africa only six HIV-1B NFLG genomes have been fully characterized, in contrast to HIV-1C NFLG genomes (LANL accessed 2016 August 30; Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997; Loxton *et al.*, 2005; Rousseau *et al.*, 2006; Wilkinson *et al.*, 2015; Junqueira and de Matos Almeida, 2016; Magiorkinis *et al.*, 2016).

In 2014, HIV-1 BC recombinants were reported circulating in South Africa (Jacobs *et al.*, 2014). In 2015, HIV-1B strain closely related to the ancient strains from the USA was also reported occurring in the heterosexual population, which indicates epidemic crossover (Middelkoop *et al.*, 2014; Wilkinson *et al.*, 2015). Considering the circulation of HIV-1 recombinant forms, there is now an increasing need for continuous HIV-1 subtype surveillance, as different subtypes can influence drug resistance and disease progression. This could be achieved through NFLG sequencing (Middelkoop *et al.*, 2014; Wilkinson *et al.*, 2015). Understanding HIV-1 diversity in South Africa could play an important role in monitoring the HIV-1 epidemic and in developing prevention and treatment strategies (Aldrich and Hemelaar, 2012; Wilkinson *et al.*, 2015).

1.5 Phylogenetic analysis of HIV

1.5.1 Concepts of molecular evolution

The science of evolution was first described in the 1800s when naturalists became fully aware that species have changed but were not sure of what changed and where the change originated. The reconstruction of evolutionary history of organisms presented in the form of a phylogenetic tree has been the main drive of all biologists ever since the time of Charles Darwin (Ayala and Fitch, 1997). The main cause behind evolution is due to mutational changes of genes. Mutational changes can occur when there is a nucleotide change caused by either a substitution, insertions or a deletion (Hartl and Clark, 1997). A phylogenetic tree consists of various components, which are interdependent of each other. The node is connected to the branches and the branch base represents the distance between the nodes (Nei and Kumar, 2000; Vandamme, 2003).

1.5.2 Multiple alignments

When raw sequences are obtained, it is important that sequences be checked for potential contaminants. In the absence of contamination, sequences need to be checked for errors and assembled into contiguous fragments (contigs). Multiple sequences that are similar has to be aligned in such a way that homologous sites appear in the same column (Siepel *et al.*, 1995; Chenna *et al.*, 2003). Sequences differ in length and gaps in some position are used to achieve alignment. Alignments are required for other analyses such as to demonstrate sequence similarity within a group of family sequences (Chenna *et al.*, 2003). Software programs that are commonly used to do alignments are Clustal W (Thompson, Higgins and Gibson, 1994; Larkin *et al.*, 2007). MAFT v7 (http://mafft.cbrc.jp/alignment/server/). Clustal X program (Chenna *et al.*, 2003; Larkin *et al.*, 2007).

1.5.3 Nucleotide substitution models

The fundamental evolution concept of DNA sequences is nucleotide substitution (transition and transversions) (Graur and Li, 1999). Scientist has developed models that take into consideration different parameters, such as base frequencies, in order to efficiently study the dynamics of nucleotide substitution (Graur and Li, 1999). The most commonly used models till date are Jukes-Cantor (Juke *et al.*, 1969), the more advanced General Time Reversible model (Rodriguez *et al.*, 1990), Kimura 2 parameter (Ng *et al.*, 2013), Felsenstein 81 (Felsenstein, 1985) and Hasegawa-Kishino-Yano HKY85 (Takahashi and Nei, 2000). There are many reviews available on the various models. Examples in the literature are Felsenstein, 1985; Saitou and Nei, 1987; Steel and Penney, 2000; Takahashi and Nei, 2000; Society, 2010; Civetta, Ostapchuk and Nwali, 2016; Yoshida and Nei, 2016.

1.5.4 Neighbour joining (NJ) trees

Neighbour Joining (NJ) is a tree building method that is based on a minimum principal of evolution (Saitou and Nei, 1987). NJ is a clustering algorithm that combines speed and accuracy. This method, at each stage of taxon clustering, uses a minimum evolution principal, but does not examine all possible topologies. The neighbours (in this instance HIV sequences) are defined as two taxa, in an unrooted tree that are connected to a single node. The algorithm tree produced begins with a star tree, which is produced under the impression that there are no clustering of taxa (Saitou and Nei, 1987; Nei and Kumar, 2000).

1.5.5 Maximum Likelihood Trees

For Maximum Likelihood (ML) approach to tree inference, the tree selected is the tree that gives the highest probability that produces the input sequence alignment (Takahashi and Nei, 2000). For every single tree, the likelihood of producing input alignment of each site is usually calculated by probabilities of summing every possible ancestral state. The likelihood at each site is the product of likelihood for the full tree. ML trees optimize the fit between the tree and the data as it searches for the topology that gives rise to the input dataset (Steel and Penney, 2000).

CONTENT

Chapter 2

Methodology

2.1 Introduction

The study was aimed to characterize NFLG sequences from identified HIV-1 B strains from the 1980s, representing the early HIV-1 epidemic in South Africa. We characterized the viral strains through NFLG amplification, sequencing and detailed phylogenetic analysis. This chapter explains the methodologies that were used to approach the study objectives. **Figure 10** illustrates the workflow breakdown of the methods used throughout this project.

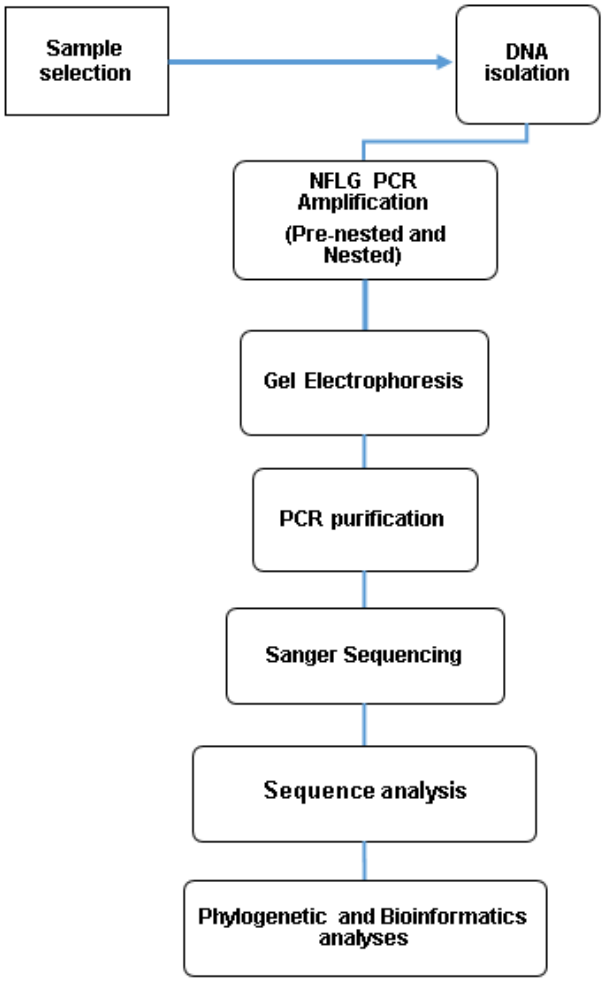


Figure 10: Flow chart illustrates the methodologies used for near full-length genome (NFLG) characterisation of HIV-1B cohort from South Africa.

2.2 Reagents and equipment

The list of chemical reagents, equipment, and software application that were used during the course of this study are listed in this chapter. In **Table B - D** shows the commercial product, equipment and software packages frequently used to perform the experiment.

Table B: List of chemical and commercial products used in the study.

Commercial products and kits used	**Application/Methods**	**Supplying Company**	**Catalogue number**
QIAamp DNA Blood Mini Kit	DNA extraction	Qiagen, Germany	51106
KAPAHiFi Ready-mix (2X) Hot start	PCR experiment	KAPA Biosystem, USA	A 1260
Nuclease free water	All experiment	Qiagen, Germany	145045078
Agarose	Gel electrophoresis	Lonza, USA	#D1 - LE
6x Blue Orange Loading Dye	Loading PCR amplicons	Promega, United States of America	G 1881
QIAquick PCR Purification Kit	PCR amplicons clean-up	Qiagen, , Germany	28 106
BigDye™ Terminator cycle sequence ready Kit	Sequencing PCR	Applied BioSystems, USA	4 337 035
5x Sequencing Buffer	Sequencing PCR	Applied BioSystems, USA	4 305 603
BigDye XTerminator Purification Kit	Sequencing clean-up kit	Applied BioSystem, USA	4 374 408

Table C: Equipment used for sample analysis

Piece of Equipment	Application/ Methods	Supplier	Location
GeneAmp PCR System 9700 thermal cycler	PCR experiment	Applied BioSystems	USA
Nanodrop™ ND 1000	Spectrophotometric measurement of DNA or RNA	Nanodrop Technologies Inc.	USA
GeneAmp® 9700 PCR system thermal cycler	PCR experiment	Applied Biosystems	USA
UV-ITEC Prochem Gel Dock System	Gel visualization and imaging	Whitehead Scientific	South Africa
Eppendorf Centrifuge 5424R	Liquid and solid separation	Eppendorf	Germany
ABI prism® 3130XL automated DNA genetic analyzer	Sequencing	Applied Biosystems	USA

Table D: Software packages and online tools used for sequence analysis

Software package	References and/or licensed companies
Sequencher v 5	Gene Codes Corporation,, USA
ClustalW v 2.1	Thompson *et al.*, 1997
MEGA v 5.0	Tamura *et al.*, 2011
MAFFT v.7	http://mafft.cbrc.jp/alignment/server/ (accessed on 2016/10/31)
BioEdit	http://www.mbio.ncsu.edu/BioEdit/bioedit.html (accessed on 2016/11/01)
jpHMM	Spang *et al.*, 2002 (accessed on 2016/10/31)
REGA v 3.0 HIV subtyping tool	de Oliveira *et al.*, 2005 (accessed on 2016/10/25)

v -version

2.3 Ethical considerations

The study received ethical approval from the Human Research Ethics Committee (HREC) at the Faculty of Medicine and Health Sciences, Stellenbosch University (Tygerberg Campus) under a main study entitled "Tracking the molecular epidemiology and resistance pattern of HIV-1 in South Africa", which is been renewed annually. The ethics approval letter is attached in **Appendix A,** ethics reference number N15/08/071.

2.4 Study population and sample selection

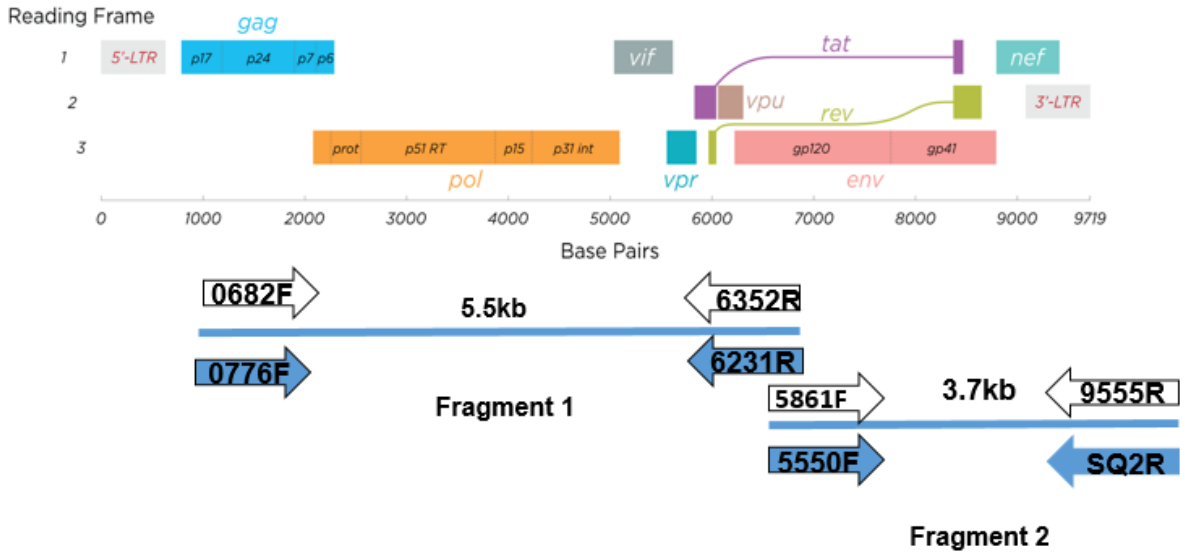
Since 1984 genomic DNA extracted from HIV-1 patients were obtained at the Tygerberg Academic Hospital in the Western Cape. In this study, we selected five previously identified HIV-1 B samples from South Africa, collected for further characterisation. All five samples included in this study were collected during the 1980s from homosexual males. The samples were partially sequenced in the *env* gene by Prof. Susan Engelbrecht (Engelbrecht, 1992). Sequences were co-cultured with donor peripheral blood mononuclear cells (PBMCs) obtained from healthy HIV negative individuals. High molecular weight (hmw) genomic DNA was sequenced from virus-infected cell cultures and stored at -20^{o}C. The current study used the same DNA from the earlier study (Engelbrecht *et al.*, 1995). The patient demographics are summarised in **Table E.**

Table E: Patient demographic

No	Sample identity	Specimen date	Year of birth	Gender	Risk factors	Origin of sample
1	R68	15/09/1987	1961	Male	Homosexual	Western Cape
2	R526	22/09/1987	1955	Male	Bi-sexual	Western Cape
3	R605	01/11/1985	1956	Male	Homosexual	Western Cape
4	R459	02/11/1987	1954	Male	Bi-sexual	Western Cape
5	R1296	30/10/1987	1938	Male	Homosexual	Western Cape

2.5 Polymerase chain reaction (PCR)

HIV-1 NFLG amplification was performed in two overlapping fragments, (Figure 11). The first fragment (F1) is from *gag* to *vpu* position relative to HXB2 (0776 > 6231) approximately 5.5kb in length. The second fragment (F2) starts from the *vpu to the3'LTR* position relative to HXB2 (5550 > 9719) approximately 3.7kb in length. **Figure 11** shows a schematic diagram of the amplification strategy as outlined by (Grossmann, Nowak and Neogi, 2015).



(Grossmann, Nowak and Neogi, 2015)

Figure 11: The amplification strategy. The diagram shows amplification for whole genome. Full genome amplification was achieved by amplifying DNA in two fragments. Four primers were used for each fragment utilising pre-nested and nested PCR. Fragment 1 consisted of genes starting from gag to vpu and was approximately 5.5Kb in length. The second fragment of genes starting from vpr until the 3'LTR and was approximately 3.7kb in length. The following paragraphs below describe each fragment in more detail.

2.5.1 Fragment 1 (F1) *Gag-Vpu*

Amplification of the *gag-vpu* gene fragment was achieved through two rounds of PCR using the high fidelity KAPA HiFi Hot Start Ready Mix (2 x) (KAPA Biosystems, USA). Reactions were performed in a final volume of 50µl. Primers for first round (0682F and 6352R) and second round PCR (0776F and 6231R) were used at an initial concentration of 10µm. **Table F** indicates the reaction composition for both the pre-nested and nested PCR master mixes.

Table F: Master Mix used for the amplification of the first fragment of HIV-1

Reagents	**Stock concentration**	**Stock Volume**	**Final concentration**	**Volume**
KAPA Hi-Fi Hot Start Ready-mix	2X	25ml	1X	25.0µl
Forward Primer (0682F/0776F)	10µM	2.0ml	300nM	1.5µl
Reverse Primer (6352R/6231R)	10µM	2.0ml	300nM	1.5µl
Template DNA	variable	variable	5ng	5µl
PCR Grade H_2O				17.0µl
Total volume				**50.0µl**

The cycling conditions used for F1 are indicated in **Table G.** The same cycling conditions were used for both first and second round PCR reactions. A GeneAmp PCR System 9700 thermal cycler (Perkin Elmer, Boston, USA) was used to perform the PCR reactions.

Table G: Cycling parameters of F1 (first and second round PCR)

Cycling conditions	**Cycle**	**Temperature in ºC**	**Time**
Initial Denaturation 1X	1	95	5 minutes
Denaturation	30	98	20 seconds
Annealing		65	15 seconds
Elongation		72	3 minutes
Final elongation	1	72	5 minutes
Reactions were stored at 4ºC until used.			

2.5.2 Fragment 2 (F2) *Vpu*- 3'*LTR*

The amplification reactions were performed using the high fidelity KAPA HiFi Hot Start Ready Mix (2X) (KAPA Biosystems, USA) Reactions were performed in a total volume of 50µl. Primers for pre-nested (5550F and 9555R) and nested PCR (5861F and SQ2R) were used at an initial concentration of 10µM. The preparation of the master mix for F2 was identical to F1, refer to **Table F** for the master mix composition. Cycling conditions for F2 are indicated in **Table H**. The same cycling conditions were retained for both pre-nested and nested PCR reactions. A GeneAmp PCR System 9600 thermal cycler (Perkin Elmer, Boston, MA, USA) was used to perform the PCR reactions.

Table H: Cycling parameters of second fragment (first and second round PCR)

Cycling conditions	**Cycles**	**Temperature in ºC**	**Time**
Initial Denaturation	1	95	5 minutes
Denaturation	30	98	20 seconds
Annealing		65	15seconds
Elongation		72	2 minutes
Final elongation	1	72	5 minutes
Reactions were stored at 4ºC until used.			

Table I shows primers name and sequences obtained from (Grossmann, Nowak and Neogi, 2015). and SQ2R was obtained from (Rousseau *et al.*, 2006). Primer 0440R is an in house primer designed in order to get a more specific amplification for F2. Human Immunodeficiency Virus Subtype B strain (HXB2) was selected because this virus is the most commonly used reference strain for many different kinds of functional studies, nucleotide position to determine position of the primers.

Table I. Primer used for PCR amplification both of fragment 1 and fragment 2

Primer name	**Sequence (5'- 3')**	**HXB2 position**
0682F	TCTCTCGACGCAGGACTCGGCTTGCTG	0682→0708
0776F	CTAGAAGGAGAGAGAGATGGGTGCGAG	0776→0800
6352R	GGTACCCCATAATAGACTGTRACCCACAA	6352→6324
6231R	CTCTCATTGCCACTGTCTTCTGCTC	6231→6207
5550F	AGARGAYAGATGGAACAAGCCCCAG	5550→5574
5861F	TGGAAGCATCCRGGAAGTCAGCCT	5861→5884
9555R	TCTACCTAGAGAGACCCAGTACA	9555→9533
SQ2R	TAGAGCACTCAAGGCAAGCTTTATTGAGGCTTA	9202→9221
0440R	CCAGAGCTCACCTAGCACCATCCAAAGGTCAGTGGG	9238→9201

2.6 Agarose gel electrophoresis

Agarose gel electrophoresis separates DNA fragments according to their size. The separation of DNA molecules happens by applying an electric field to a gel matrix. A 0.8% agarose gel was prepared by completely dissolving 0.80g of molecular grade agarose (Invitrogen, USA) in 100 ml of 1 X Tris-Acetate-EDTA (TAE) buffer, which was heated in the microwave oven. The 1 X TAE buffer was prepared from a stock of already prepared 50 X TAE buffer. Molten agarose was stained with 10µl of GR green [New England Biolabs, United Kingdom (UK)]. Subsequently, poured into a plastic gel electrophoresis tank and allowed to set at room temperature. Five microliter (5µl) of each PCR product, mixed with 1µl of 6 x loading dye (Thermoscientific, USA), was loaded into gel lanes and run alongside negative and positive controls and a 1kb molecular weight DNA marker (Promega, USA). DNA was resolved by electrophoresis at 75 Volts for 50 minutes. The resolved DNA fragments were visualized with the Alliance Chemilluminscence and Fluorescence gel imaging system (UVItec,

UK). The expected fragment sizes for the first fragment, F1 *(gag-vpu) and* second fragment, F2, *(vpu-3'LTR)* were 5.5kb, and 3.5kb respectively.

2.7 DNA purification

Visualized positive PCR products were purified. Briefly, DNA purification was achieved using two separate purification methods. The first method was performed using Direct PCR purification with the QIAquick PCR purification kit (Qiagen, Germany). The second purification method was carried out using QIAquick gel extraction kit (Qiagen, Germany) for PCR product that had multiple bands visible on the gel. Purification was done according to the manufacturer's instructions, both methods are described below.

2.7.1 QIAquick PCR purification method (Direct PCR purification)

The first method was performed using the QIAquick PCR purification kit (Qiagen, Germany). Briefly, in a clean 1.5ml centrifuge tube, 115µl of solubilisation buffer (buffer QG) was added to 5 volumes of PCR product. The resulting mixture was transferred into a mini elute column and centrifuged at 13000 revolutions per minute (rpm) for 1 minute (Eppendorf Centrifuge 5424R, Germany), allowing the binding of viral DNA to the gel matrix column. The flow-through was discarded. A volume of 750µl of wash buffer (buffer PE) was added to the column and centrifuged at 13000 rpm for 1 minute. The flow through was discarded in order to remove residual buffer. The column was centrifuged at 13000 rpm for 1 minute and the collection tube discarded. The column was placed in a clean 1.5ml centrifuge tube and 45µl of elution buffer (buffer EB) was added to the centre of the column. The column assembly was incubated at room temperature for 3 minutes and centrifuged at 13000 rpm for 1 minute. The column was discarded and the purified DNA product was confirmed by resolving a 2µl aliquot of the product on a 0.8% agarose gel-by-gel electrophoresis.

2.7.2. QIAquick gel extraction method

The second purification method used was agarose gel extraction with the QIAquick gel extraction kit (Qiagen, Germany). Briefly, 45µl of the PCR product was loaded on a 0.8% agarose gel and subjected to electrophoresis. A sterile blade was used to excise the band of interest under a dual intensity ultraviolet transilluminator (Uvp Inc, USA). The excised band was placed in a nuclease free 1.5ml centrifuge tube and weighed (Mettler instrumente, Switzerland). Subsequently, 300µl of buffer QG was added to the excised gel thereafter incubated at 50°C with pulse vortexing (Heidolph, Germany), for complete dissolution of the gel. A 100µl of aliquot isopropanol was added to the dissolved gel to

precipitate the DNA. The mixture was added to a spin column which was placed into a collection tube and centrifuged at 13 000 rpm for 1 minute in the Eppendorf centrifuge 5424R (Eppendorf, Germany). Once the DNA bound to the membrane of spin column, the flow through was discarded. An extra centrifugation step was performed at 13 000 rpm to remove residual buffer. The column was placed in a clean micro-centrifuge tube and 45µl of Buffer EB was added to the centre of the column to elute the DNA. The column assembly was incubated at room temperature for 3 minutes and centrifuged at 13 000 rpm for 1 minute. The purified products were confirmed by resolving a 2µl aliquot of the product on a 0.8% agarose gel-by-gel electrophoresis.

2.8 Sequencing of near full-length genome (NFLG) of HIV-1

Purified DNA products concentration were determined on the ™ND1000 Spectrophotometer (Nanodrop technologies, USA). Primers used for the sequencing are indicated in Appendix 2. Sequencing primers were diluted to a final concentration of 5µmol with nuclease free water (Qiagen Germany). For each sequencing reaction mixture, the composition was as follows: 1.0µl of the BigDye® X Terminator Enzyme mix, 3.0µl of 5X reaction buffer, 1.0 µl of each primer at a concentration of 5µmol, 1.0µl of purified DNA and 4.0µl nuclease free water to a final volume of 10µl. **Table J** illustrates the cycling parameters of the PCR sequencing assay. Whole genome sequencing primers consisted of a combination of in house designed primers and primers described by Rousseau *et al*., 2006; see attached in **(Appendix 2)**. The sequencing PCR products were purified using the BigDye® XTerminator Purification kit and analysed on an ABI 3130XL genetic analyser (Applied Biosystems, USA). The ABI 3130XL is a polymer-based capillary electrophoresis sequencer connected to a computer with sequencing analysis software (Applied Biosystems, USA) and captures raw data trace files. Raw data trace files in the form of sequences chromatograms were retrieved from the 3130XL ABI Prism genetic analyser and imported into Sequencher version 5.0 (Gene Codes Corporation, USA). Sequences with poor quality ends were removed from overlapping nucleotide sequences to improve the quality of the sequences and assembled into a single contig file. Assembled contig files were screened for ambiguities, edited, exported and saved in fasta format as a text file (.txt).

Table J: sequencing PCR cycle parameters

Cycling conditions	Cycles	Temperature in °C	Time
Initial denaturation	1	95	60 seconds
Denaturation	25	95	60 seconds
Annealing		55	7 seconds
Elongation		60	4 minutes
Reactions were stored in 4°C until further used			

2.9 Sequence quality control

The nucleotide sequences were verified for stop codons, insertion and deletions using an online quality control tool on the HIV LANL. HIV BLAST was used to find sequences similar to patient sequences. Thereafter, patients sequences were characterized and subtyped using the Jumping Profile of Hidden Markov Model (jpHMM), Recombinant Identification Program (RIP) v 3.0and REGA v3.0 phylogenetic analysis (http://www.hiv.lanl.gov/content/sequence/QC//index.html).

2.10 Characterization of sequences using online HIV subtyping tools

Three online subtyping tool were used for data analyses namely; REGA version 3.0, jumping profile Hidden Markov Model (jpHMM), and Recombinant Identification Program (RIP) v 3.0. All three subtyping tool make use of different algorithms. Results obtained from each tool gives an idea of what subtype our query sequences. Results obtained from each subtyping tool were subjected to phylogenetic analyses to confirm the reliability of the patient sequence subtype.

2.10.1 REGA version 3.0 subtyping analysis

REGA (http://bioafrica.mrc.ac.za:8080/rega-genotype-3.0.2/hiv/typingtool#/) is an online genetic subtyping tool that makes effective use of bootscanning methods and phylogenetic analysis to detect recombination and subtype of specific gene fragments. The tool is able to analyses 1000 sequences at any given time and has been previously used to characterize HIV-1 subtypes. REGA is also accessible from the LANL webpage (http://www.hiv.lanl.gov/). (de Olivera *et al.*, 2005)

2.10.2 Jumping Profile Hidden Markov Model Analysis (jpHMM)

The jumping profile Hidden Markov Model (jpHMM) (http://jphmn.gobics.de/) is a method that compares query nucleic acid or protein sequences to a family of sequences in a jumping alignment approach (Spang *et al.*, 2002). A single query sequence is therefore aligned and compared to different sequences of a single alignment. This approach simplifies the identification of recombinant events as breakpoints of recombinant isolates will be compared to different sequences in the alignment (Zhang *et al.*, 2006; Schultz *et al.*, 2009).

2.10.3 Recombinant Identification Program (RIP) version 3.0

The Recombinant Identification Program (RIP) (http://www.hiv.lanl.gov) is an online program that provides a summary of information to describe large sets of sequences. This online program makes use of two program options. The first option is an “informative mode” which when activated counts only informative positions. The informative position is the position at which at least one subtypes consensus sequence differs from the other in related to a central or reference position. The second option relates to the handling of gaps. Gaps in RIP can either be “squeeze” or “strip.” RIP will only squeeze gaps that are not in position where query or all subtype consensus sequences are represented by a gap. Gaps can be strip gaps which means that it ignores position at which the query or any subtype consensus sequences is represented by a gap (Siepel *et al.*, 1995).

2.10.4 Multiple alignments of query and reference sequences

HIV-1 subtypes complete genome reference sequences were retrieved from the LANL (http://www.hiv.lanl.gov) database. Alignments of the reference sequences were done using Clustal X (Larkin *et al.*, 2007). All multiple alignments were converted to Mega format (. meg). The Genbank Accession number and family subtype was used to label all the reference sequences.

2.10.5 Model test

Model test is an important test performed on aligned sequences when you intend to draw a Maximum likelihood (ML) phylogenetic tree. In order to know which test should be used to generate the phylogenetic tree a model test must be carried out on the aligned query sequences and the reference sequences. This test is carried out on Molecular Evolutionary Genetics Analysis (MEGA) v6.06. Lowest BIC scores were considered as the best model test (Tamura *et al.*, 2013). The lowest BIC for each samples were the General Time Reversible (GTR) with a discrete Gamma distribution by assuming that a certain site is evolutionary invariable (I).

2.10.6 Construction of ML trees

MEGA version 6.06 was used to analyze converted MEGA files (.meg) (Tamura *et al.*, 2013). Model tests were carried out using MEGA v6.06. Lowest Bayesian Information Criterion (BIC) scores are considered to describe the best substitution pattern. Oftentimes, the model with the lowest BIC when constructing a ML tree are General Time Reverse Gamma distributed with invariant sites (GTR+G+I), General Time Reverse Gamma distributed (GTR+G), The accuracy of the generated tree was validated by bootstrap analysis with a total of 1000 bootstrap replicates for each dataset.

Content

Chapter 3

Results

3.1 Introduction

This chapter focuses on NFLG analysis of the HIV-1B samples and sequences obtained for my study. PCR amplification were performed on selected samples. The samples were amplified in two separate fragments (F1 and F2) and yielded approximately 9kb. Subsequently, amplicons were sequenced and generated sequences were analyzed and subtyped using various online subtyping tool. Finally, phylogenetic analysis was performed on the complete sequence dataset for full characterization of HIV-1B NFLG sequences.

3.2 DNA quantification and PCR amplification of the HIV-1 genome

Table L shows selected samples DNA concentration as determined on the Nanodrop. The concentration ranges from 156 µg/µl.and the purity ranges from 302 µg/µl. Subsequently, NFLG PCR amplification were performed on the selected samples for the study. **Table L** below gives a brief overview samples amplified and amplicons that were sequenced.

3.3 PCR amplification

Table K: Nanodrop concentration, PCR amplified and sequenced amplicons

			Amplified PCR product		Sanger sequenced amplicons	
No	**Sample identity**	**DNA Concentration (µg/µl)**	**F1 (Gag-Vpu)**	**F2 (Vpu-3'LTR)**	**F1 (Gag-Vpu)**	**F2 (Vpu-3'LTR)**
1	R1296	156	Yes	Yes	Yes	Yes
2	R526	187	Yes	Yes	Yes	Yes
3	R68	179	Yes	Yes	Yes	Yes
4	R459	275	Yes	Yes	Yes	Yes
5	R605	302	Yes	Yes	Yes	Yes

Positive amplicons were confirmed to be the right size (~5.5kb) with reference to the positive control. Absence of a band in the negative control well confirmed the absence of contamination **Figure 12A** shows ~5.5 kb PCR amplicons and **Figure 12b** shows~3.7 kb PCR amplicons.

A

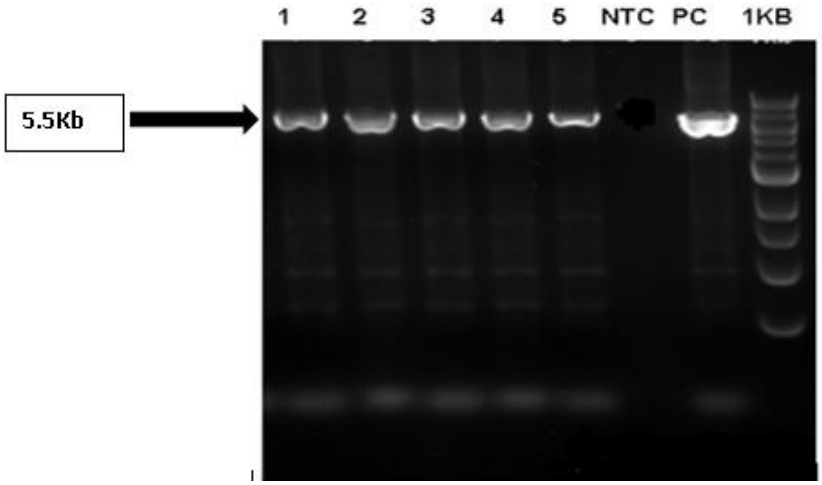
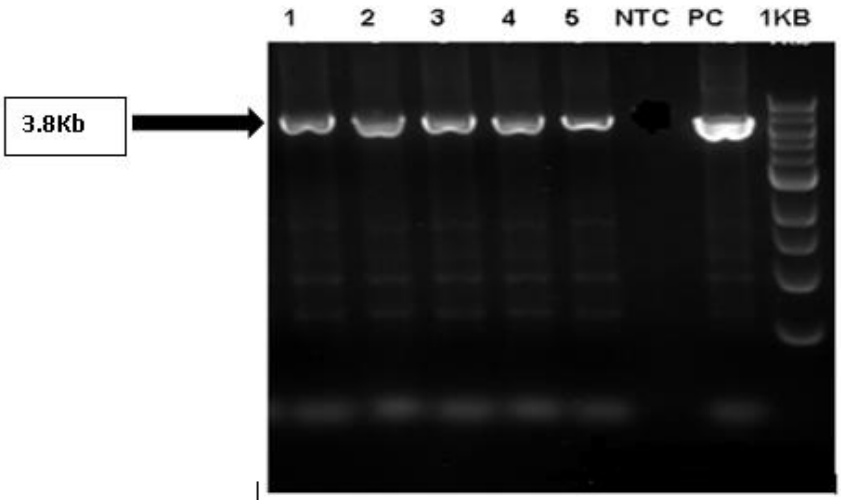


Figure 12: A 0.8% agarose gel of F1. The gel shows a visible, specific band of the F1 (~5.5kb) in size for five patient amplicons, negative template control (NTC), Positive control (PC) and 1KB ladder. The amplified ~5.5kb fragments were purified.

12 B



In Figure 12 B: **A 0.8% agarose gel electrophoresis**. This gel shows a visible, specific amplification of the F2 (3.7kb) in size for five patient amplicons, negative template control (NTC), Positive control (PC) and 1KB ladder. Amplification was achieved using two sets of primers in two rounds of PCRs. The amplified 3.7kb fragments were purified.

3.4 DNA sequencing

Five amplicons with both fragments were amplified through PCR and subsequently sequenced using the ABI automated sequencer. Direct sequencing of the second fragment consumed more time and was more challenging due to the variability in the *env* gene. **Figure 13** shows an example of the sequence chromatogram of the *env* gene region. The *env* gene is the least conserved region of the HIV-1 genome. It is always constantly under attack by the host immune system because it helps with the fusion of virus membrane to the host cell membrane. In response, the virus make use of its glycoprotein to protect itself by mutating the *env* gene. The read lengths of the five NFLG sequences were at an average of approximately 8797bps. See Tables **3.3 – 3.6** for detailed information on gene annotation for each sample. Sequences with quality below 75% were removed and sequences were trimmed from edges to remove low-confidence base calls. Gene specific primers were designed to sequence and fill in the missing gaps to give readable sequences. The sequencing primers are listed in **Appendix 2.**

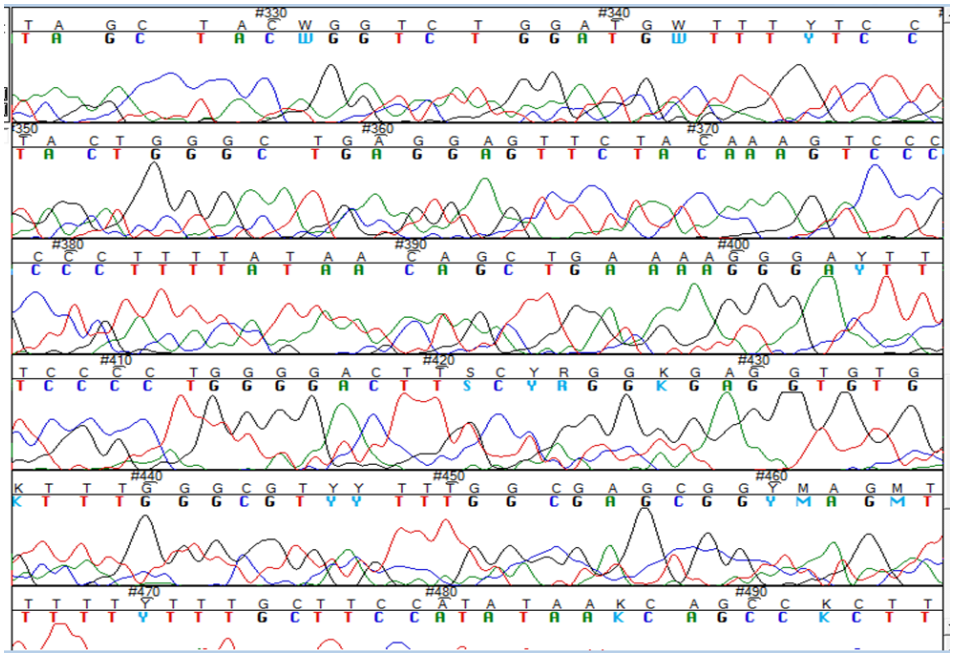


Figure 13: A snapshot of sequencer chromatogram of the *env* gene. The F2, which includes the *env* gene gave problematic sequences and was hard to read. New gene specific primers were designed to help compensate for the problem.

3.5 Sequence analyses

3.5.1 Assembling of NFLG

Sequences were checked for quality control and quality scores between 75 and 90% were obtained to form NFLG contigs. Sequences were trimmed at the edges to remove low-confidence base calls and were manually edited using the Sequencher v5 software. Sequences with high quality were manually assembly. Thereafter, they were put together to form contigs. **Figure 14** shows snapshot from Sequencher v5, which shows how the NFLG were completely sequenced using both forward and reverse reads.

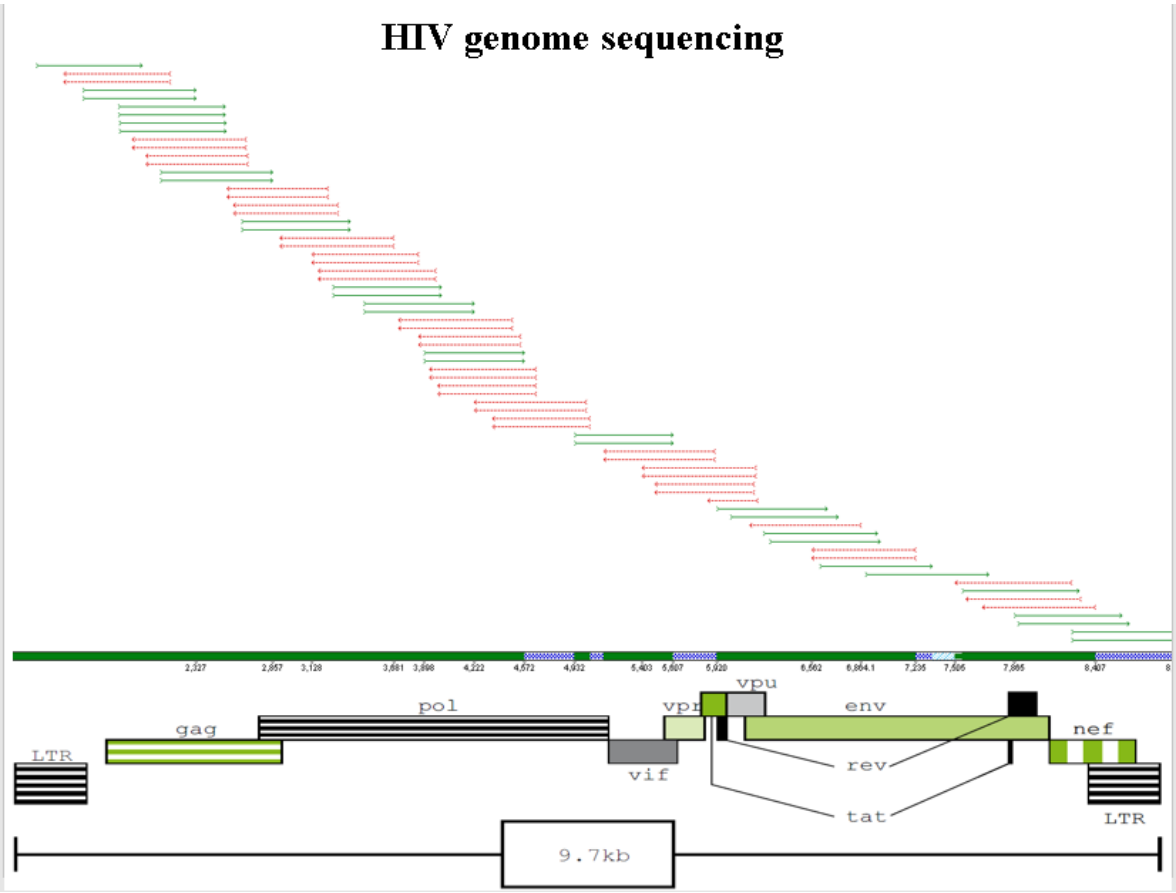


Figure 14: A snap shot of HIV-1 genome sequencing from Sequencher v5. The green arrows represent the forward primers; the red arrows represent the reverse primers. The sequences were mapped against the HXB2 reference strain.

3.5.2 Annotation of genes

In order to have an indication of the start and end point of each viral gene, gene annotation were done for all NFLG sequences relative to the HXB2 reference strain. Tables **M – Q** The HIV/SIV sequence locator on the HIV database was used for annotation of each viral gene. All the query viral genes are similar in length compared to HXB2 sequences from LANL. **ZA|85|R68** has premature stop codons in the *env* gene. The *env* encodes for the Envelope protein located at the outer shell membrane of the virus. In response to the host immune attack the *env* gene mutates to avoid attack from host immune. The stop codons are at the following positions *env* 370, 383 with no frameshift detection. **ZA|86|R526** has three frameshifts at the following positions assesory gene *vif* 173, *vpr* 2, and *nef* 144 with no premature stop codons. **ZA|86|R605** also has no premature stop codon, but with 5 frameshift detected at the respective positions *gag* 417, *pol* 3, *env* 81, *env* 728 and *nef* 126 **ZA|85|R459** has 3 frameshifts at the respective positions *gag* 1, *rev* 79, and *env* 775 with no premature stop codons. **ZA|87|R1296** has 3 frameshifts at the following position in the *env* 194, *env* 466 and *nef* 142 with no premature stop codons. It is important to know that the HIV/SIV sequence locator might give misleading results in cases where the query sequences have an insertion or a deletion. Sequence alignment should be done in order to be certain with the sequence integrity.

Sample **ZA|85|R68** sequence starting position relative to HXB2 complete genome starts from the *gag* at position 886 and ends at the *nef* at position 8825. The following online subtyping tool; jpHMM, REGA and RIP were used in F**igure 15** for sample **ZA|85|R68** sequence analysis's in order to fully characterize the sequence.

Table L: Nucleotide position on the HIV genome relative to sample ZA.85.R68

Region	Nucleotide position relative to HXB2 genome (Start – End)
Gag	886 → 2292
Pol	2085 → 5096
Vif	5041 → 5619
Vpr	5559 → 5850
Tat 1	5831 → 6045
Rev1	5970 → 6045
Vpu	6062 → 6310
Env	6225 → 8795
Tat 2	8379 → 8469
Rev 2	8379 → 8653
Nef	8797 → 8825

Sample **ZA|87|R526** sequence starting position relative to HXB2 complete genome starts from the *gag* position 886 and ends at the *nef* position 8825. The following online subtyping tool; jpHMM, REGA and RIP were used in **Figure 16** for **ZA|87|R526** sequence analyses in order to fully characterize the sequence.

Table M: Nucleotide position on the HIV genome relative to sample ZA.86.R526

Region	Nucleotide position relative to HXB2 genome (Start → End)
Gag	849 → 2292
Pol	2085 → 5096
Vif	5041 → 5619
Vpr	5559 → 5850
Tat 1	5831 → 6045
Rev1	5970 → 6045
Vpu	6062 → 6310
Env	6225 → 8795
Tat 2	8379 → 8469
Rev 2	8379 → 8653
Nef	8797 → 9175

Sample **ZA|85|R605** sequence starting position relative to HXB2 complete genome starts from the *gag* position 849 and ends at the *nef* position 9175. The following online subtyping tool; jpHMM, REGA and RIP were used in **Figure 17** for the **ZA|87|R605** sequence analyses in order to fully characterize the HIV-1 genome.

Table N: Nucleotide position on the HIV genome relative to sample ZA.86.R605

Region	Nucleotide position relative to HXB2 genome (Start › End)
Gag	849 → 2292
Pol	2085 → 5096
Vif	5041 → 5619
Vpr	5559 → 5850
Tat 1	5831 → 6045
Rev1	5970 → 6045
Vpu	6062 → 6310
Env	6225 → 8795
Tat 2	8379 → 8469
Rev 2	8379 → 8653
Nef	8797 → 9175

Sample **ZA|85|R459** sequence starting position relative to HXB2 complete genome starts from the *gag* position 849 and ends at the *rev* position 8539. The following online subtyping tool; jpHMM, REGA and RIP were used in **Figure 18** for the **ZA|87|R459** sequence analyses in order to fully characterize the HIV-1 genome.

Table O: Nucleotide position on the HIV genome relative to sample ZA 85.R459

Region	Nucleotide position relative to HXB2 genome (Start →End)
Gag	849 → 2292
Pol	2085 → 5096
Vif	5041 → 5619
Vpr	5559 → 5850
Tat 1	5831 → 6045
Rev1	5970 → 6045
Vpu	6062 → 6310
Env	6225 → 8539
Tat 2	8379 → 8469
Rev 2	8379 → 8539

Sample **ZA|85|R1296** sequence starting position relative to HXB2 complete genome starts from the *gag* position 1054 and ends at the *3'LTR* position 9719. The following online subtyping tool; jpHMM, REGA and RIP were used in **Figure 19** for the **ZA|87|R1296** sequence analyses in order to fully characterize the HIV-1 genome.

Table P: Nucleotide position on the HIV genome relative to sample ZA. 87. R1296

Region	Nucleotide position relative to HXB2 genome (Start – End)
Gag	1054 → 2292
Pol	2085 → 5096
Vif	5041 → 5619
Vpr	5559 → 5850
Tat 1	5831 → 6045
Rev1	5970 → 6045
Vpu	6062 → 6310
Env	6225 → 8795
Tat 2	8379 → 8469
Rev 2	8379 → 8653
Nef	8797 → 9417
LTR3	9086 → 9719

3.6 Online subtyping analyses of HIV-1

In order to fully characterize each of the sample sequence online subtyping tool such as RIP, jpHMM and REGA were used to comprehensively analyse samples selected for the purpose of the study. **Figure 15 – Figure 19** represent a visual description of each sequence analyses and the genome subtype. All the online subtyping tool used in the following; **Figure 15, 16, 18 and 19** confirms the NFLG as a pure subtype B except for **ZA|85|R605.** In **Figure 17 ZA|85|R605** is confirmed as a HIV-1 BD recombinant with the following break points; *gag* = D, *pol* = B; *env* = B. There is a discrepancy with the *nef 3'LTR* – RIP and REGA says B, while jpHMM says K, A1. In order to have a comprehensive analysis of the sample we drew a phylogenetic tree for each regions of the genome.

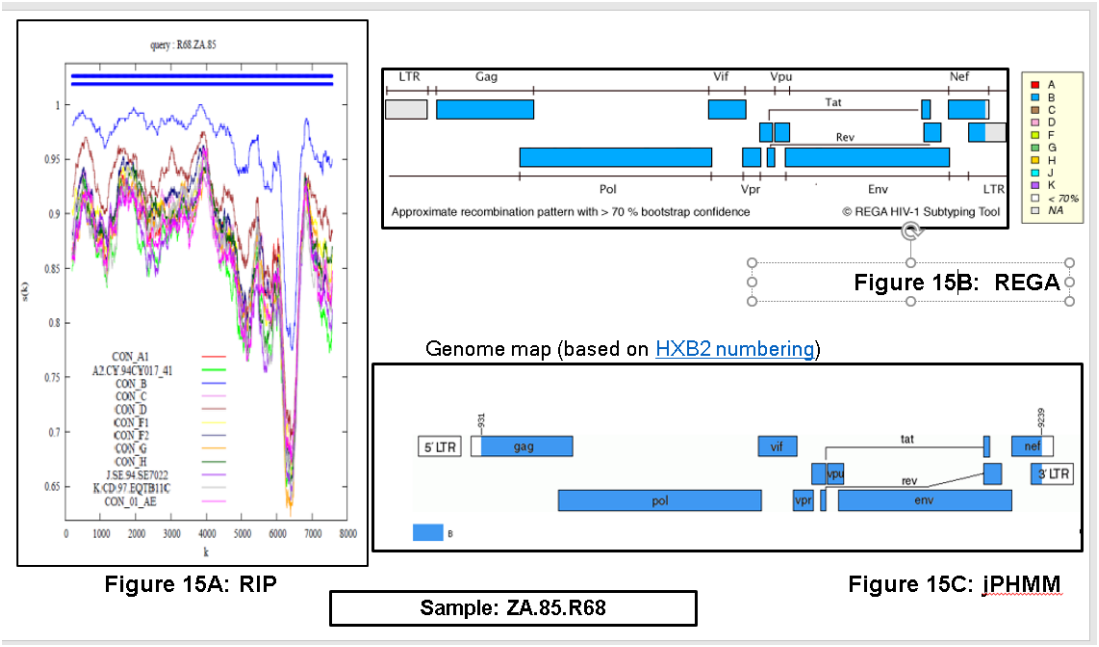


Figure 15: Analysis of sample ZA|85|R68 using three online HIV-1 subtyping tool 15A: RIP; 15B: REGA; and 15C: jpHMM. All the three HIV-1 online subtyping tool indicates *gag* = B, *pol* = B; *env* = B; *nef* = B and 3*'LTR*= B. In order to validate the online subtyping tool result, we drew a phylogenetic tree for NFLG of HIV-1B.

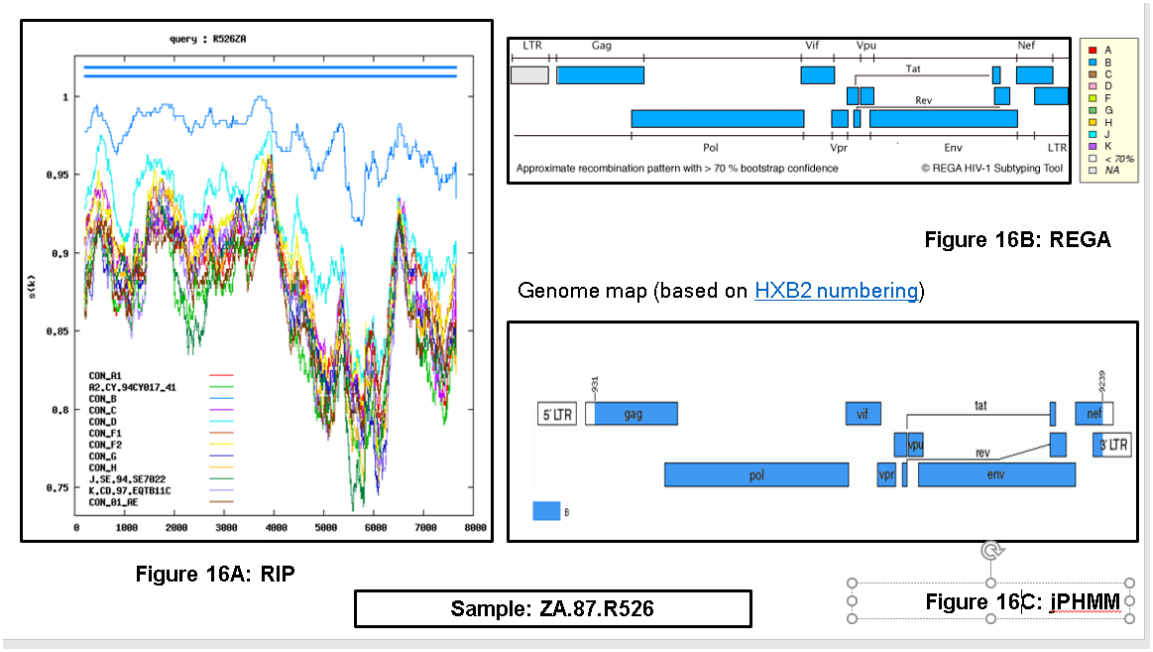


Figure 16: Analysis of sample ZA|87|R526 using three online HIV-1 subtyping tool. 16A: RIP; 16B: REGA; and 16C: jpHMM. All the three HIV-1 online subtyping tool indicates gag = B, pol = B; env = B; nef = B and 3'LTR= B. In order to validate the online subtyping tool result, we drew a phylogenetic tree for NFLG of HIV-1B.

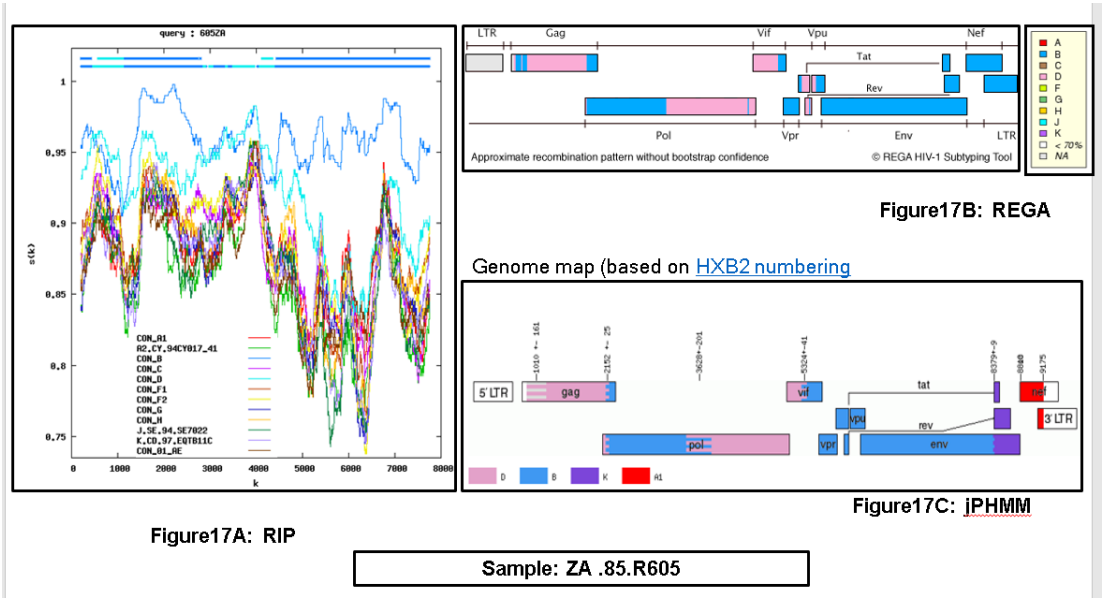


Figure 17: Shows analysis of sample ZA|85|R605 using three online HIV-1 subtyping tool. 17A: RIP; 17B: REGA; and 17C: jpHMM. *gag* = D, *pol* = b; *env* = b; Discrepancy with the *nef 3'LTR* – RIP and REGA says B, while jpHMM says K, A1. In order to have a comprehensive analysis of the sample we drew a phylogenetic tree for each regions of the genome. The breakpoints are listed in **Table O**.

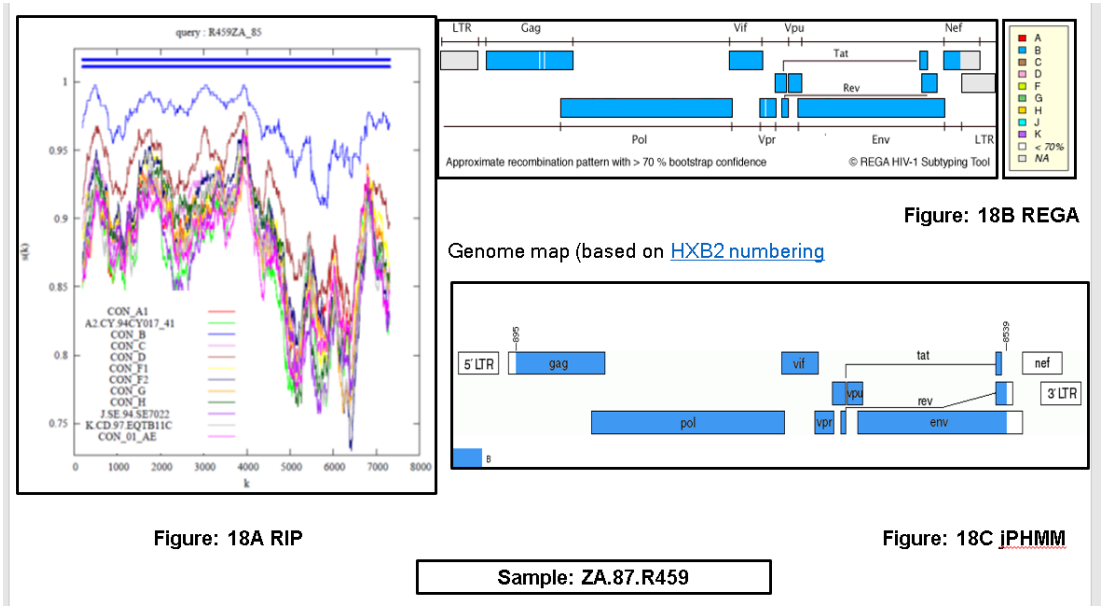


Figure 18: Analysis of sample ZA|87|R459 using three online HIV-1 subtyping tool. 3.5A: RIP; **3.5B**: REGA; and **3.5C**: jpHMM. All the three HIV-1 online subtyping tool indicates *gag* = B, *pol* = B; *env* = B; *nef* = B and 3*'LTR*= B. In order to validate the online subtyping tool result, a phylogenetic tree was drawn for NFLG of HIV-1B.

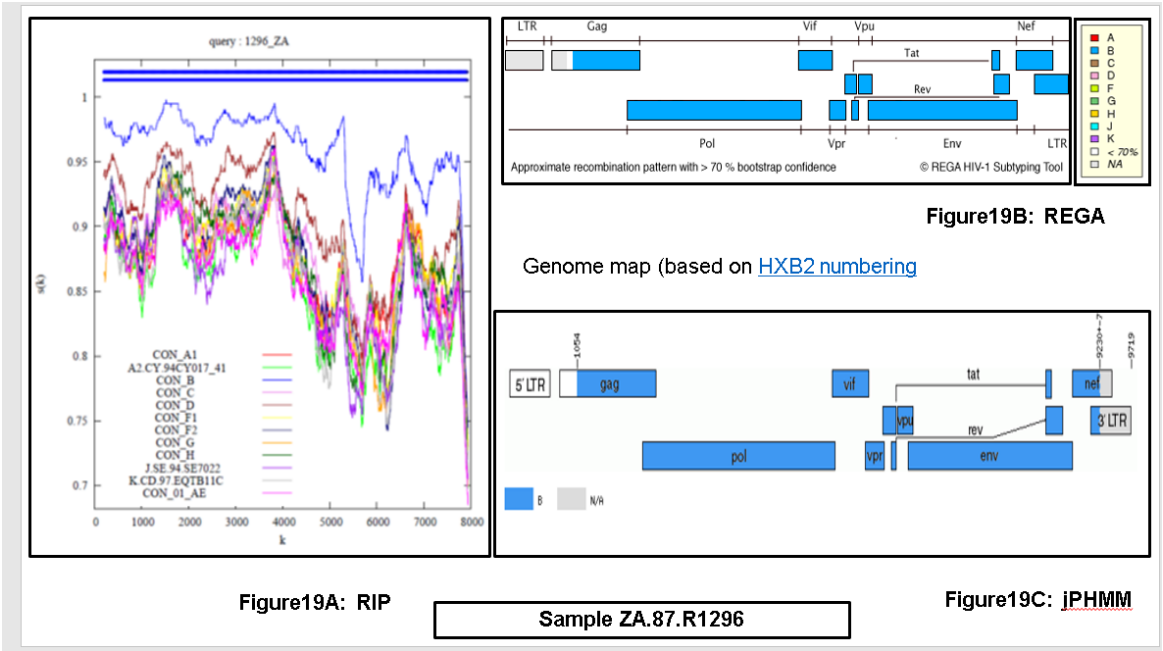


Figure 19: Analysis of sample ZA|87|R1296 using three online HIV-1 subtyping tool. 19A: RIP; **19B**: REGA; and **19C**: jpHMM. All the three HIV-1 online subtyping tool indicates *gag* = B, *pol* = B; *env* = B; *nef* = B and 3*'LTR*= B. In order to validate the online subtyping tool result, we drew a phylogenetic tree for NFLG of HIV-1B.

3.7 Phylogenetic analysis

In a phylogenetic tree, epidemiologically related sequences that share a common or recent ancestor should cluster together. In order to define the phylogenetic connection amongst the NFLG sequences of R68, R1296, R526, R459 and R605 we performed a total of nine multiple alignments. In **Figure 20,** the ML phylogenetic tree shows the NFLG analyses of patient sequences in comparison to NFLG group M subtype reference without recombinants. The second NFLG phylogenetic tree in **Figure 21** phylogenetic tree shows the NFLG analyses of patient sequences in comparison to other subtype B sequences. Reference sequences selection criteria was based on HIV BLAST results with closely related sequences to the patient sequences. Additional reference sequences were obtained from Worobey *et al.*, 2016.

The NFLG ML phylogenetic analyses of patient sequences indicate that reference viral sequences are related with significant statistical bootstrap value of 99%. Patient sequences clusters together with subtype B reference sequences at the top of the phylogenetic tree. Sequence of patient **ZA 86. R526** clusters with USA strain from 1979 with GenBank accession number KJ 704791. Sequence of patient **ZA.85. R68** clusters with the USA strain from 1986 with GenBank accession number AY 835771. Sequence of patient **ZA.85. R459** clusters with the USA strain from 1986 with GenBank accession number M93259 through a BLAST search. Sequence of patient **ZA.87. R1296** clusters with the Netherlands strain from 1986 with GenBank accession number AY 970947 through a BLAST search. Additional HIVB reference sequences mostly from the New York and San Francisco in the USA were obtained from Worobey *et al.*, 2016. These sequences probably cluster together as HIV-1B epidemic in South Africa is thought to have come from contacts in the USA and Europe. In addition, HIV-1B clade sequences from Haiti, Trinidad and Tobago obtained from (LANL accessed 30th of August 2016). Sequence of patient **ZA.86. R605** is a HIV-1 subtype B outlier, which support the presence of BD recombinant at the start of the epidemic.

3.8 NFLG analysis

Figure 21 presents ML phylogenetic evaluation of patient NFLG sequences with HIV-1B and HIV-1D reference sequences obtained from the LANL HIV database. These sequences are amongst the oldest HIV-1B NFLG identified. HIV-1B reference sequences were obtained from South Africa, USA, France, Haiti and Trinidad and Tobago. Reference sequences selection criteria were based on the role that each country played with the spread of HIV-1B. We also included five HIV-1D sequences from South Africa due to the identification of the novel HIV-BD recombinant sequence from patient **ZA**|**85**|**R605**. Four of the five patient samples are marked at the tree node with a rectangle and the sequence from sample **ZA**|**87**|**R1296** is marked by a triangle. The sequence from sample **ZA**|**85**|**R605** is a subtype B outliner. This sample clusters in between reference sequence from the USA, with the year and ascension number 2005|B|US|2005|01226, and reference sequence 1983|D|CD|193|ELI with a 100% bootstrap value. The sequence from sample **ZA**|**87**|**R526** clusters in between sequences from South Africa, with the year and ascension number B|ZA|1985|R84|FJ647145, with a 100% bootstrap value. The sequence from sample **ZA**|**87**|**R68** clusters in between reference sequence with the year and ascension number B|FR|1983|HB2-LAI-IIB-BRU and 1983|PZ|04053R84 with a 100% bootstrap value as well. The sequence from sample **ZA**|**87**|**R1296** clusters in between reference sequence from South Africa and the USA with the year and ascension number B|ZA|2000|TV047|KJ94866 and B|US|1983|SF2 LAV2 ARV2. The sequence from sample **ZA**|**87**|**R459** clusters in between reference sequences from New York in the USA. This reference sequence is amongst the oldest HIV-1B strains identified by Worobey *et al*., 2016, with the year and ascension number |B|US|1979|4NY79. Subgenomic phylogenetic analysis illustrates that the sequence from sample **ZA**|**85**|**R605** is a HIV-1 BD recombinant, **Figure 24 – 27**.

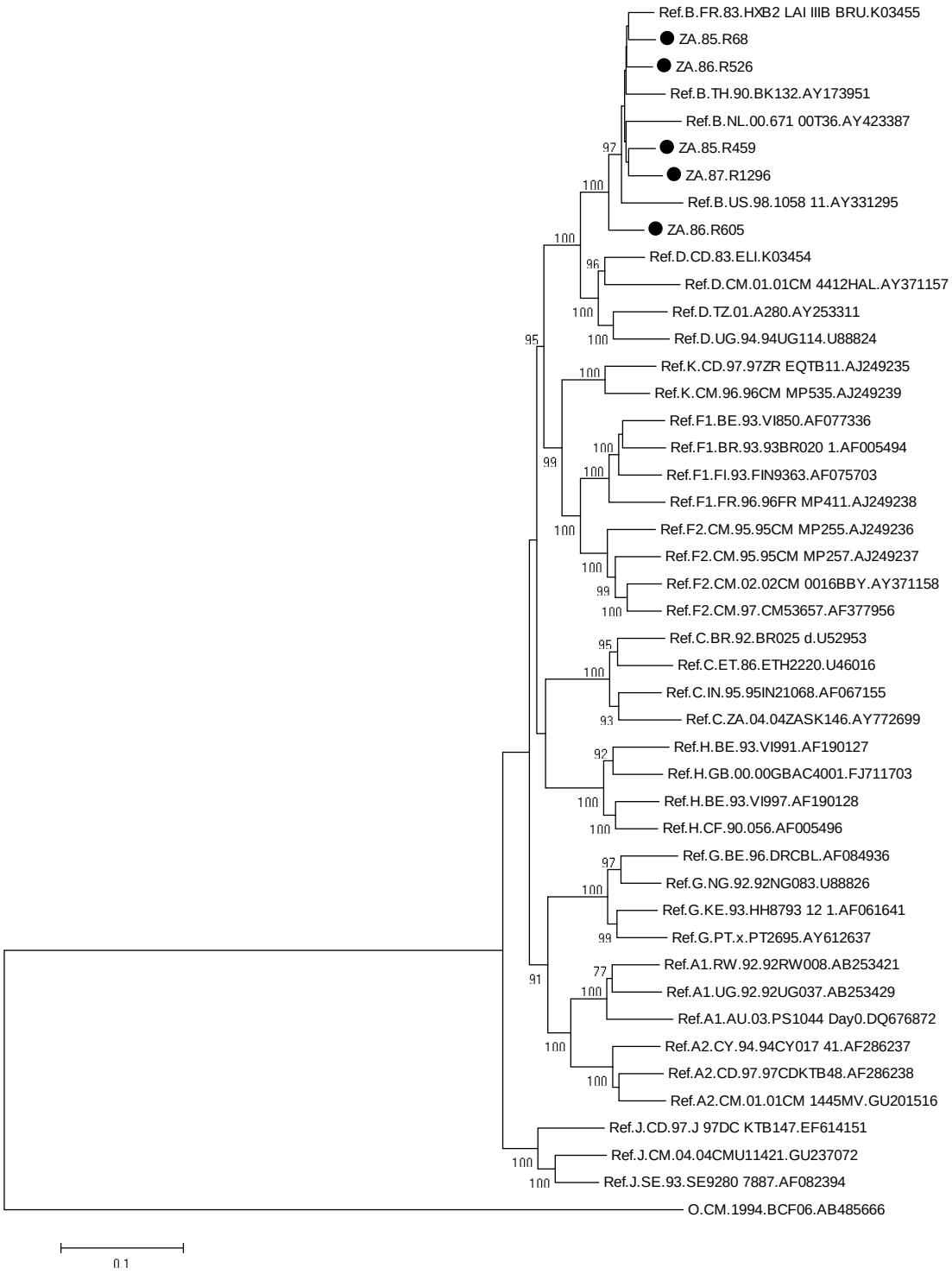


Figure 20: A ML phylogenetic tree of HIV-1 NFLG sequences. The analysis involved 38 nucleotide sequences. Bootstrap values greater than 70% are shown at the main nodes. Position according to HXB2 (849 – 9719). Horizontal scale 0.1 scale was used for the branch lengths.

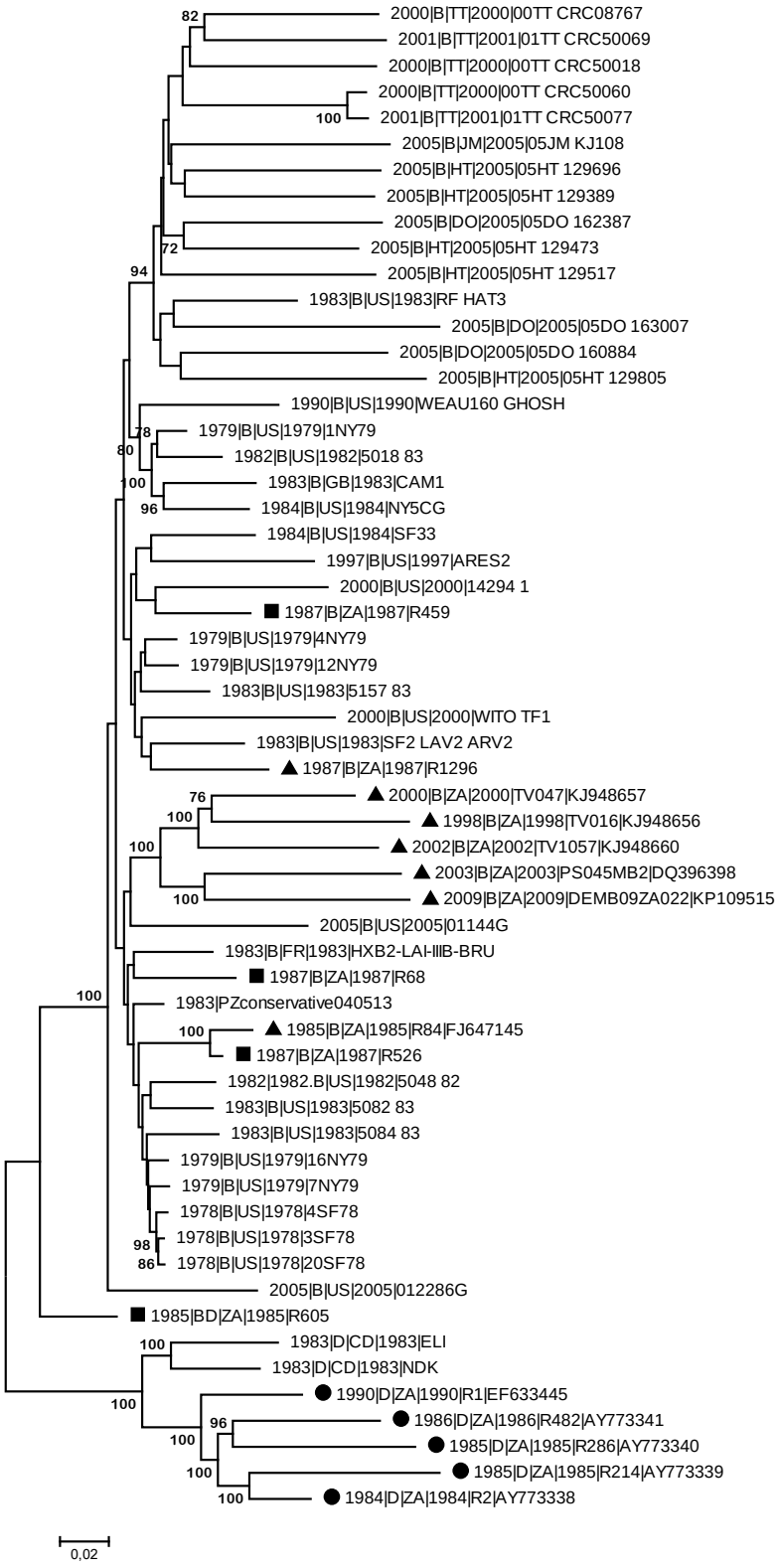


Figure 21: A ML phylogenetic tree of HIV-1 NFLG sequences. Bootstrap values greater than 70% are shown at the main nodes. P**osition** according to HXB2 (849 – 9719) Horizontal scale 0.02 scale was used for the branch lengths.

3.9 Phylogenetic sub-genomic fragment analysis of sample ZA.85.R605

In order to have a comprehensive analysis the sequence of sample **ZA|85|R605.** We drew a ML phylogenetic tree for each break point on the HIV-1 genome with respect to jpHMM. **Figure 22 – 25** show the phylogenetic analysis according to the subtype break points as indicated in **Figure 17**. The analysis confirms the **ZA|85|R605** is a recombinant HIV-1 BD.

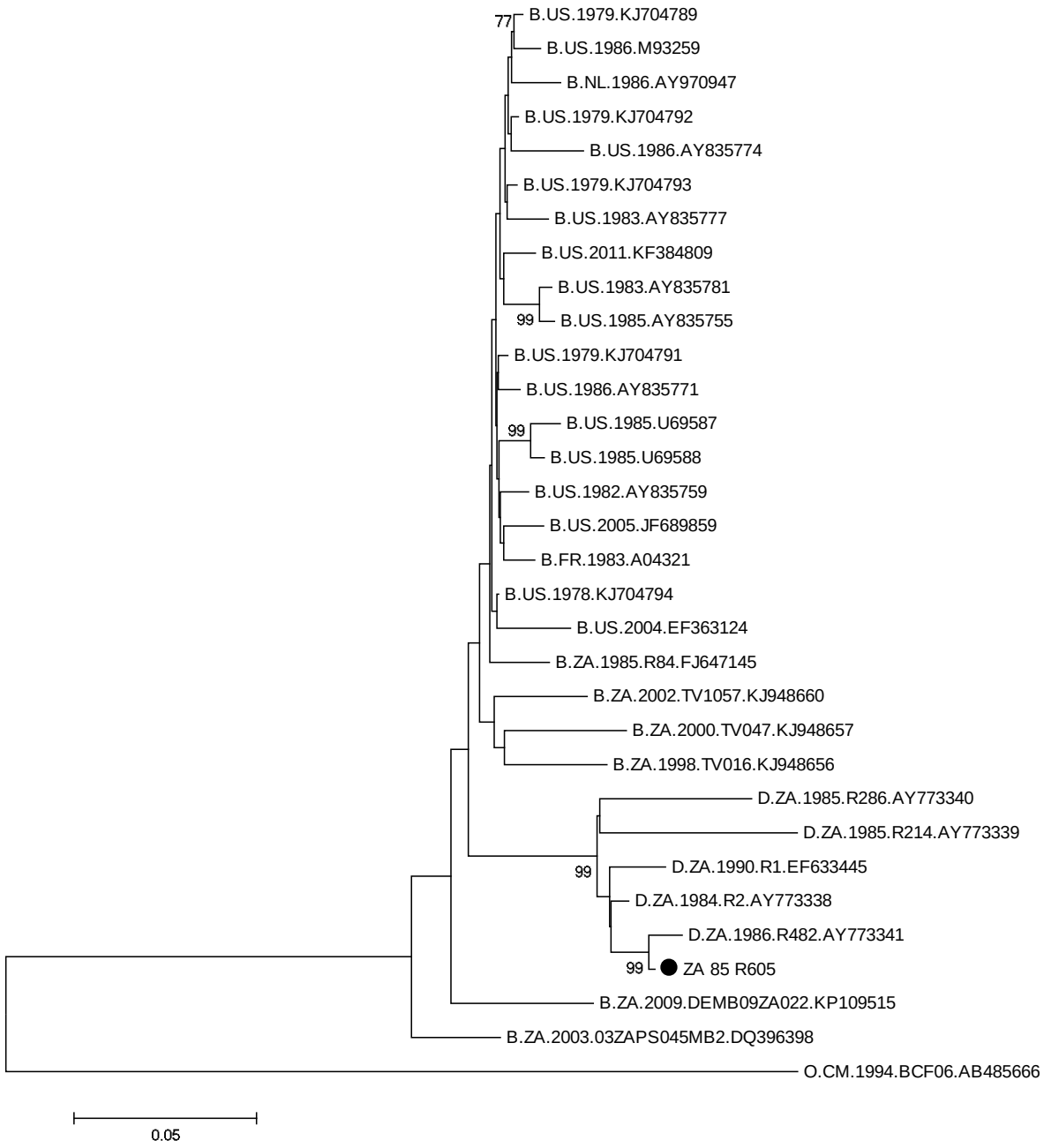


Figure 22: A ML phylogenetic tree of the sequence of sample ZA|85|R605. Bootstrap values greater than 70% are shown at the main nodes. Position according to HXB2 (1010 – 2152). Horizontal scale 0.05 scale was used for the branch lengths.

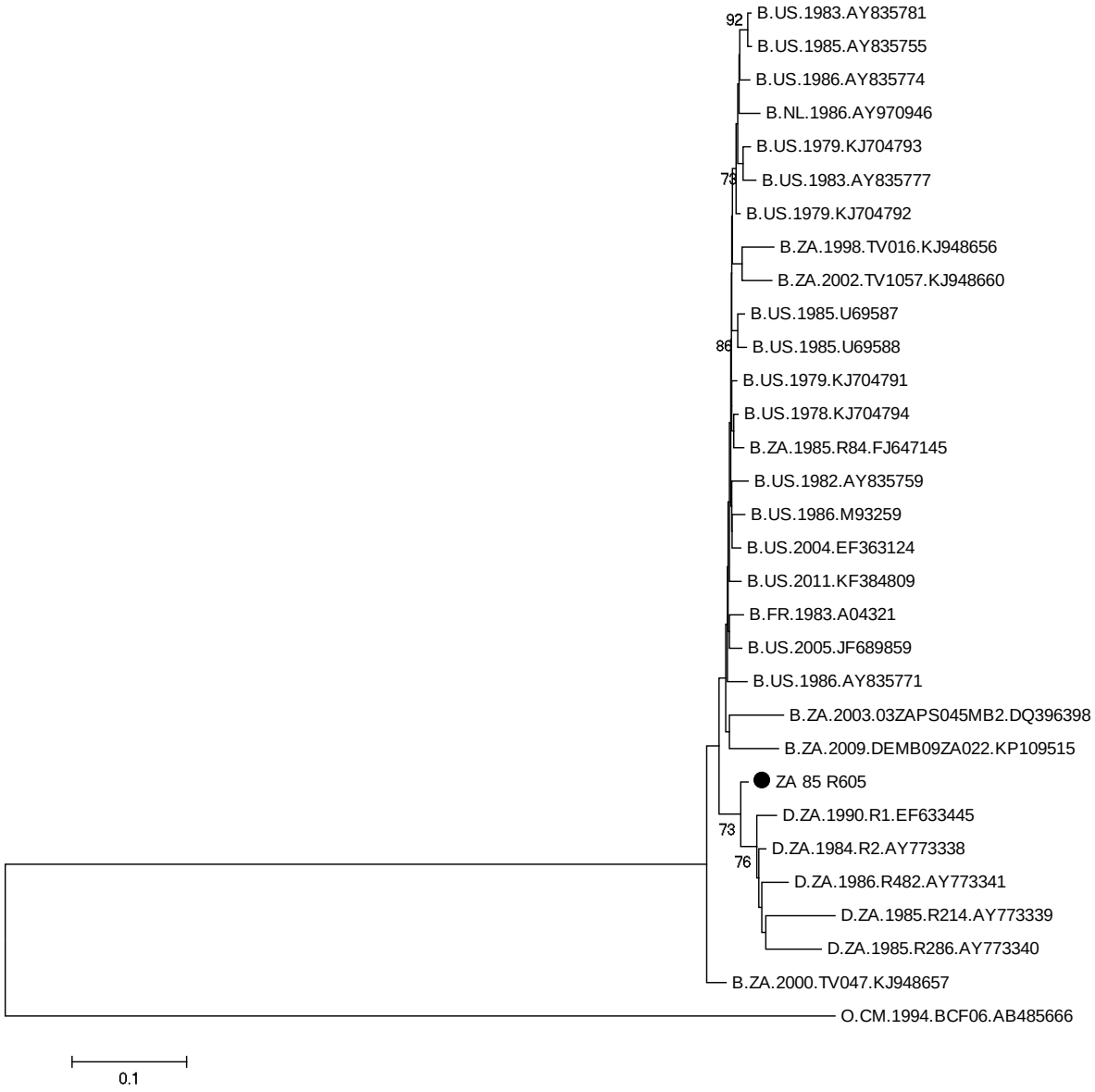


Figure 23: A ML phylogenetic tree sequence of sequence ZA|85|R605 subtype B Bootstrap values greater than 70% are shown at the main nodes. Position according to HXB2 (2152 – 3626) Horizontal scale 0.05 scale was used for the branch lengths

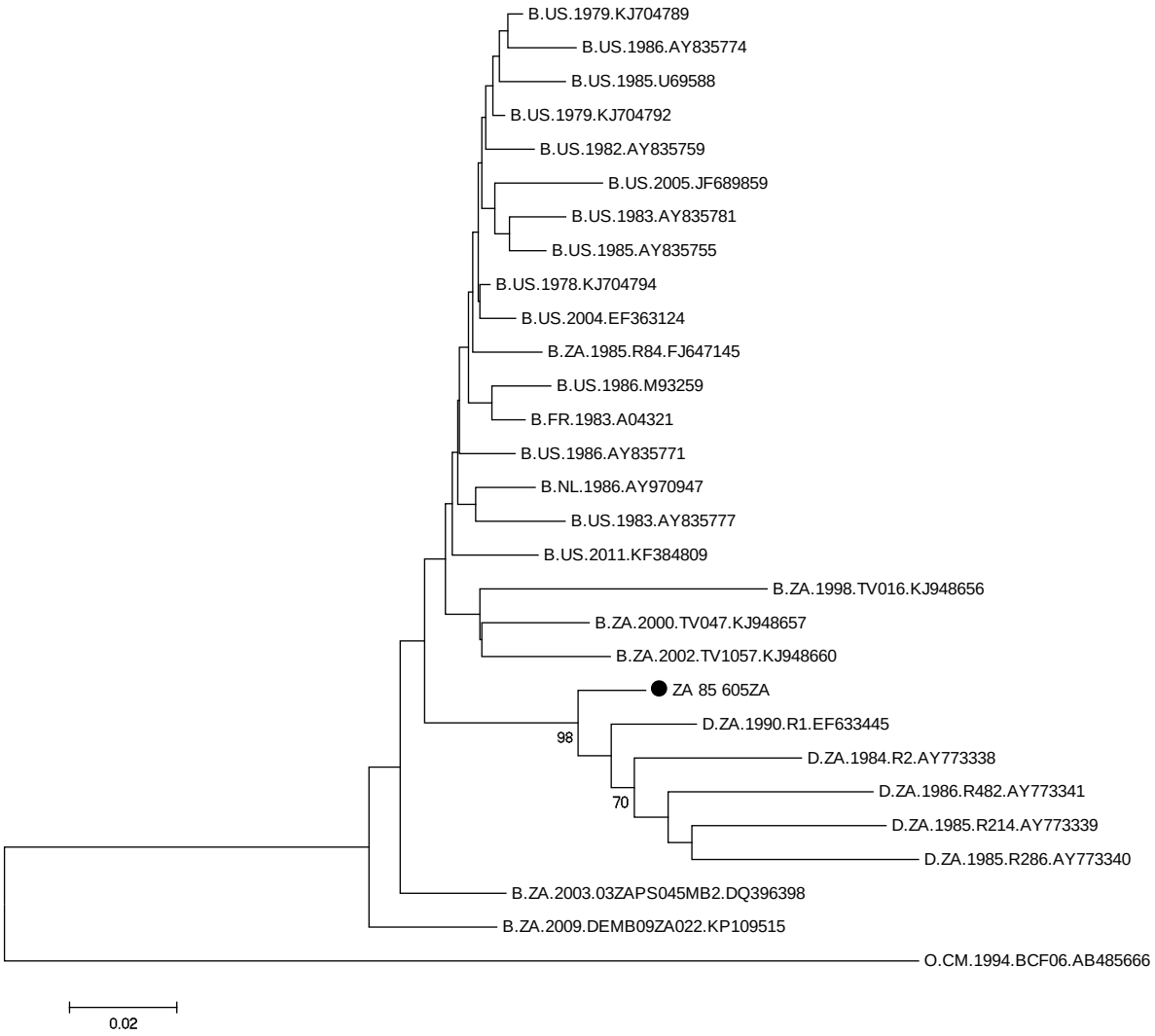


Figure 24: A ML phylogenetic tree of sequence ZA|85|R605 subtype D. Bootstrap values greater than 70% are shown at the main nodes. Position according to HXB2 (3626 – 5324) DNA sequences. Horizontal scale 0.05 scale was used for the branch lengths

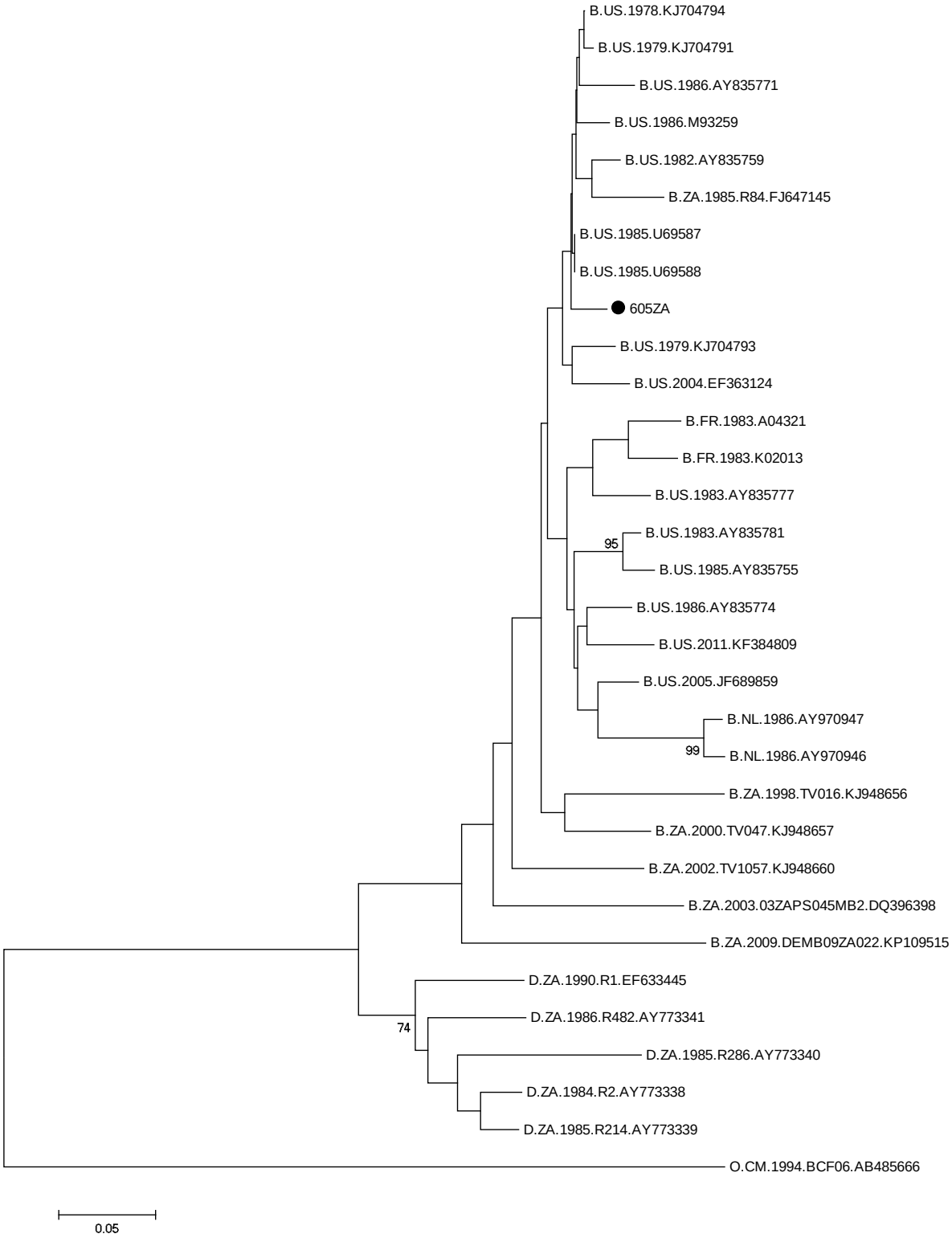


Figure 25: A ML phylogenetic tree of the sequence of sample ZA|85|R605. Bootstrap values greater than 70% are shown at the main nodes. Position according to (HXB2 5324 – 8379). Horizontal scale 0.05 scale was used for the branch lengths.

Content

Chapter 4

Discussion

4.1 Introduction

In this section, we discuss the significance of the data generated from this study. The five HIV-1B samples used for this study were previously genotyped in the *env* region from patients infected through homosexual transmission during the 1980s (Engelbrecht *et al.*, 1995). The samples were fully sequenced and characterized during this study. Four of the five HIV-1 NFLG sequences were characterized as pure HIV-1B. One identified as an HIV-1BD recombinant. An evolutionary phylogenetic approach was used in the analyses of the sequences with reference sequences obtained from the HIV database, LANL. The HIV-1BD recombinant identified is a novel recombinant strain and represents the first and only NFLG HIV-1BD sequence analysed thus far. It is the only NFLG HIV-1BD recombinant in the world.

4.2 The HIV-1 epidemic in South Africa

Phylogenetic analyses of the NFLG sequences in **Figure 21** shows that all the archival samples clusters with the South African, North American and European strains. The results generated from this study are in agreement with studies that suggest that the HIV-1B was first introduced into South Africa by the Europeans and the North Americans (Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997). Homosexual flight stewards, international tourists and migrants from the European and North American countries were most likely responsible for the introduction of HIV-1B epidemic into South Africa (Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997), refer to **Figure 2.**

In South Africa two independent epidemics have been described (Engelbrecht *et al.*, 1995; Van Harmelen *et al.*, 1997; Loxton *et al.*, 2005). Scientific reports indicate that in the mid-1980s HIV-1B was predominantly circulating amongst the white MSM and HIV-1C amongst the heterosexual population (Van Harmelen *et al.*, 1997). Online subtyping programmes used for characterization of the archival samples shows, that the findings from this study further strengthen the hypothesis of the presence of HIV-1B amongst the homosexual risk group at the beginning of the epidemic in South Africa, refer to **Figure 16 – 19.** Furthermore, to the best of our knowledge we are reporting for the first time HIV-1BD recombinant, at the beginning of the epidemic, refer to **Figure 17**.

Scientific reports during the early days regularly used smaller gene portions to characterize HIV-1 subtypes, which could have limited the identification of recombinant viral strains. These archival samples were stored as hmw DNA at -20°C for long-term storage purposes by Prof. Susan Engelbrecht. With the advancement of molecular techniques and the availability of archival samples, we managed to characterize a recombinant strain from the beginning of the epidemic.

Jacobs *et al* conducted a study in 2009 in which 22 out of 320 (6.9%) sequences were characterized as HIV-1B. In 2014, Jacobs *et al*., 2014 also reported the presence of HIV-1BC recombinant strains in the Western Cape Province of South Africa. Middelkoop *et al*., (2014) also reported the presence of HIV-1B and HIV-1C circulating amongst the MSM in South Africa. In 2015 HIV-1B strains that were closely related to ancient strains from the USA were reported to be circulating in the heterosexual population, which indicates epidemic crossover (Middelkoop *et al.*, 2014; Wilkinson *et al.*, 2015). All these findings further strengthen the hypothesis that the state of HIV-1 epidemic in South Africa is constantly changing.

4.3 The significance of the HIV-1B epidemic in SA and Africa

Limited information is available for HIV-1B sequences from South Africa. Only six NFLG HIV-1B sequences from South Africa are available in the LANL HIV database (Wilkinson *et al.*, 2015; https://www.hiv.lanl.gov/components/sequence/HIV/search.comp, accessed 2016 August 30). Four out of the six are from the Stellenbosch University Tygerberg Virology Division and Wilkinson *et al*., 2001, 2015, describe these strains. Rousseau *et al*., 2006, have described one of the six. Hora *et al* (https://www.hiv.lanl.gov/components/sequence/HIV/search.comp accessed 2016 August 30) described the latest sequence. In Africa, there is only one other published NFLG sequence from Gabon (Huet *et al*., 1989). The last HIV-1B NFLG was sequenced in 2009, which means that for the past seven years not much focus has been placed on NFLG of HIV-1B in both Africa and South Africa. The HIV-1B from Gabon shows no phylogenetic relationship with any of the South African HIV-1B sequences. The sequence was most closely related to the 1980s strains from the USA with GenBank ascension number AY835781 and with the Netherland strains from 1986 with GenBank ascension number U34603. Although the infection occurred in the 1980s, analyses of ancestral sequences are critical to understanding the time of origin of the early epidemic in the country (Wilkinson *et al.*, 2015).

4.4 The significance of NFLG sequences of HIV

HIV-1 NFLG sequencing provides important data that is essential for understanding viral evolution and the pathogenesis of HIV infection (Rousseau *et al.*, 2006; Grossmann, Nowak and Neogi, 2015). Previous molecular epidemiological studies have repeatedly used smaller (for example *gag*, *pol* and *env*) or partial genes to evaluate HIV-1 forms of recombination (Neogi *et al.*, 2012; Grossmann, Nowak and Neogi, 2015). Previous studies have stressed the importance of NFLG to validate major HIV-1 circulating recombinant forms (Wilkinson and Engelbrecht, 2009; Jacobs *et al.*, 2014; Wilkinson *et al.*, 2015). Therefore, the effective use of NFLG as a tool for HIV-1 subtyping is important to understand the dynamic of the HIV epidemic within different populations (Gall *et al.*, 2012; Grossmann, Nowak and Neogi, 2015; Wilkinson *et al.*, 2015). The NFLG sequencing protocol developed from this study has been used to characterize our HIV-1BD URF.

Previously developed NFLG protocols are either disadvantaged by low quantity, which is often limited to a particular HIV-1 subtype or does not amplify proviral DNA (Rousseau *et al.*, 2006; Grossmann, Nowak and Neogi, 2015). This protocol has successfully amplified four major subtypes and recombinant forms (HIV-1B, HIV-1C, 01_AE, BD and 02_AG) which are collectively responsible for more than 80% of global infections (Hemelaar, Gouws, Peter D. Ghys, *et al.*, 2011; Grossmann, Nowak and Neogi, 2015).

Over the years, there have been several approaches to develop NFLG protocols. Fang *et al.*, (1996) developed two different NFLG protocol strategies: the first protocol targeted to amplify a 9-kb fragment, whilst the second protocol targeted to amplify two overlapping 5 kb fragments. Both protocols were developed from plasma viral RNA for efficient HIV-1 NFLG cloning, using reverse transcriptase and long PCR amplification strategies. NFLG protocols were also developed by Gao *et al.*, (1996), who directly sequenced the amplicons, without first cloning.

Recent approaches with next-generation sequencing (NGS) platforms have revolutionized the field of HIV-1 NFLG sequencing and analyses. NGS provides exceptional possibilities for large-scale sequencing. NFLG protocols using NGS platforms has been described by (Archer *et al.*, 2012; Gall *et al.*, 2012; Grossmann, Nowak and Neogi, 2015). The feasibility of NGS platform in low middle-income countries (LMICs) has been limited due to cost. Currently, there are several platform options accessible for NGS, such as Ion Torrent and Illumina. It is important that we consider developing NGS protocol for complete genome amplification suitable for our setting.

4.5 The HIV-1 epidemic in South African homosexual males

The current HIV research emphasis in South Africa has been on the heterosexual and vertical transmission (Middelkoop *et al.*, 2014). There has been little information about HIV-1 subtypes circulating in the homosexual risk group (Middelkoop *et al.*, 2014). The impact of homosexual transmission patterns in South Africa may be minor, however it cannot be ignored (Middelkoop *et al.*, 2014). In the mid 80's and 90's, due to the political segregation of apartheid, inter-racial relationships were prohibited and different cultural backgrounds may have influenced the identity and activity of heterosexuals (Middelkoop *et al.*, 2014). Over the past decades changing sociopolitical landscapes and behavioral patterns have played a significant role in the spread of the epidemic (Abecasis *et al.*, 2013; Hawke *et al.*, 2013; Takebe *et al.*, 2014; Junqueira and de Matos Almeida, 2016). The change in behavioral evidence formed a link between heterosexuals and homosexuals (Middelkoop *et al.*, 2014). Recent studies have described the presence of unique inter-subtype recombinants such as BC, BF, and AC in both homosexuals and heterosexuals (Jacobs *et al.*, 2014; Middelkoop *et al.*, 2014; Wilkinson *et al.*, 2015). Wilkinson *et al* (2015) also support the theory of a link between the homosexual and heterosexual population. All these findings strengthen the theory of a potential epidemic crossover. The presence of a BF sequence and a rare F2 sequence is also a cause for concern because it shows the increasing presence of recombinant strains in South Africa (Middelkoop *et al.*, 2014). Countries in different regions of the world have full documentation of HIV-1 strains circulating amongst the homosexual population and HIV-1B prevalence has been reported in the homosexual risk groups all over (Arán-Matero *et al.*, 2011; Abecasis *et al.*, 2013; Hawke *et al.*, 2013; Takebe *et al.*, 2014; Junqueira and de Matos Almeida, 2016). HIV-1B dissemination in various parts of the country has been largely influenced by the change in sexual orientation (Junqueira and de Matos Almeida, 2016). Asia has witnessed a dramatic genotype switch as opposed to other parts of the world (Kato *et al.*, 2003; Tee *et al.*, 2005; Kondo *et al.*, 2013). To date, limited studies have been aimed at understanding the origin and spread of HIV-1B in homosexuals (Chen *et al.*, 2011; Kondo *et al.*, 2013; Ng *et al.*, 2013; Takebe *et al.*, 2014; Junqueira and de Matos Almeida, 2016).

4.6 HIV-1 subtype D in South Africa

There are in total 79 NFLG pure HIV-1D sequences; only seven (8.8%) of these are sequences from outside Africa (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.comp, accessed on the 21/11/2016). Engelbrecht *et al* (1995), from our laboratory, described five HIV-1D viruses from South Africa. Apart from these five not a lot of focus has been placed on the virus from this country (Loxton *et al.*, 2005; Jacobs *et al.*, 2007). In 1997, HIV-1D was identified in a male homosexual patient and one heterosexual patient through partial *gag* analysis (van Harmelen *et al*., 1997). Bredell *et al*., (2002) conducted a study where he identified an HIV-1 CD recombinant. The Democratic Republic of Congo (DRC) has reported a high prevalence of HIV-1D (11.5%) as compared to South Africa and many other African countries (Vidal *et al*., 2000). According to Jacobs *et al*., unpublished data from Stellenbosch University Tygerberg Virology, our laboratory has also identified the presence of possible HIV-1CD recombinant in the country. Wilkinson *et al*., (2015) characterized an NFLG URF HIV-1AD strain from South Africa. The NFLG characterization of HIV-1BD is an indication that viral diversity was present at the beginning of the epidemic.

4.7 Strengths and limitations

One of the strengths of the study was the ability to amplify the NFLG of archive samples using high efficient, modern molecular techniques. Such improvements are reflected by the successful troubleshooting and optimization of NFLG PCR reactions and the designing of gene-specific primers suitable for the NFLG sequencing. The optimized NFLG protocol can potentially be used in large-scale population-based molecular epidemiological studies. It can also be implemented in the clinical laboratory management of drug resistance genotyping with the advantage of amplifying full-length Gag-Pol for determining predictors of PI, RT and IN drug mutation, as well as genotypic co-receptor tropism testing for co-receptor antagonists. The protocol will also help fill in the gap for more NFLG sequences because partial HIV-1 sequences may underrepresent viral recombinant forms. Sequencing of the *env* gene was challenging and significant amount of time was lost. The study limitations include the scarce availability of archive ancestral samples. As the sample volumes are often limited, we have one chance to apply our NFLG protocol to obtain valuable sequence data. Thus, these stored samples need to be handle with great care. In addition, the protocol has only been optimized for proviral DNA amplification and not viral RNA.

4.8 Ongoing / Future work

This study forms the basis for continued research in our attempt to reconstruct the epidemiology and evolutionary history of HIV in South Africa. In the current study, we did not investigate epidemic crossover (for example, we did not include samples from different time points) due to time constraints and availability of funds. Future work can focus on looking at the transmission clusters and looking into how the epidemic has spread from the beginning until now. We are also planning to perform NGS on these archive amplicons to establish that the protocol is not Sanger sequence dependent but can be sequenced on both Sanger and NGS platforms. We are also currently in the process of optimizing this protocol to amplify NFLG from viral RNA.

4.9 Conclusion

This work represents the most recent analyses of archival HIV-1B sequences from the early South African epidemic. Our results highlight the importance of using HIV NFLG sequences to give a more detailed picture of early landmarks in the HIV/AIDS epidemic. This study also adds valuable information to the origin and history of HIV-1 in South Africa. To the best of our knowledge, we have characterised the first HIV-1BD recombinant strain in the world. The presence of HIV-1 BD recombinant at the beginning of the epidemic is an indication that viral recombination events was happening at the beginning of the epidemic in South Africa, but could have easily been missed as sequence analyses were often limited to small genomic regions of HIV. Comprehensive modern NFLG sequence analyses can be used as a valuable tool to map the evolutionary path of HIV in South Africa.

References

Abecasis, A. B., Wensing, A. M. J., Paraskevis, D., Vercauteren, J., Theys, K., Van de Vijver, D. A. M. C., Albert, J., Asjö, B., Balotta, C., Beshkov, D., Camacho, R. J., Clotet, B., De Gascun, C., Griskevicius, A., Grossman, Z., Hamouda, O., Horban, A., Kolupajeva, T., Korn, K., Kostrikis, L. G., Kücherer, C., Liitsola, K., Linka, M., Nielsen, C., Otelea, D., Paredes, R., Poljak, M., Puchhammer-Stöckl, E., Schmit, J.-C., Sönnerborg, A., Stanekova, D., Stanojevic, M., Struck, D., Boucher, C. A. B. and Vandamme, A.-M. (2013) 'HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics.', *Retrovirology*, 10(1), p. 7. doi: 10.1186/1742-4690-10-7.

Aldrich, C. and Hemelaar, J. (2012) 'Global HIV-1 diversity surveillance', *Trends in Molecular Medicine*, 18(12), pp. 691–692. doi: 10.1016/j.molmed.2012.06.004.

Ammann, A. J., Abrams, D., Conant, M., Chudwin, D., Cowan, M., Volberding, P., Lewis, B. and Casavant, C. (1983) 'Acquired immune dysfunction in homosexual men: Immunologic profiles', *Clinical Immunology and Immunopathology*, 27(3), pp. 315–325. doi: 10.1016/0090-1229(83)90084-3.

Arán-Matero, D., Amico, P., Arán-Fernandez, C., Gobet, B., Izazola-Licea, J. A. and Avila-Figueroa, C. (2011) 'Levels of spending and resource allocation to HIV programs and services in Latin America and the Caribbean', *PLoS ONE*, 6(7), pp. 1–9. doi: 10.1371/journal.pone.0022373.

Archer, J., Baillie, G., Watson, S. J., Kellam, P., Rambaut, A. and Robertson, D. L. (2012) 'Analysis of high-depth sequence data for studying viral diversity: a comparison of next generation sequencing platforms using Segminator II', *BMC Bioinformatics*. BioMed Central Ltd, 13(1), p. 47. doi: 10.1186/1471-2105-13-47.

Becker, M. L., De Jager, G. and Becker, W. B. (1995) 'Analysis of partial gag and env gene sequences of HIV type 1 strains from southern Africa.', *AIDS research and human retroviruses*, 11(10), pp. 1265–1267.

Beyrer, C., Baral, S. and Griensven, F. van (2012) 'Global epidemiology of HIV infection in men who have sex with men', *The Lancet*, 380(9839), pp. 367–377. doi: 10.1016/S0140-6736(12)60821-6.Global.

Beyrer, C., Sullivan, P., Sanchez, J., Baral, S. D., Collins, C., Wirtz, A. L., Altman, D., Trapence, G. and Mayer, K. (2013) 'The increase in global HIV epidemics in MSM.', *Aids*, 27(17), pp. 2665–78. doi: 10.1097/01.aids.0000432449.30239.fe.

Bobkov, A., Kazennova, E., Selimova, L., Bobkova, M., Khanina, T., Ladnaya, N., Kravchenko, A., Pokrovsky, V., Cheingsong-popov, R. and Weber, J. (1998) 'A Sudden Epidemic of HIV Type 1 among Injecting Drug Users in the Former Soviet Union : Identification of Subtype A , Subtype B , and Novel gagA / envB Recombinants', 14(8).

Cabello, M., Junqueira, D. M. and Bello, G. (2015) 'Dissemination of nonpandemic Caribbean HIV-1 subtype B clades in Latin America.', *AIDS (London, England)*, 29(4), pp. 483–92. doi: 10.1097/QAD.0000000000000552.

Cabello, M., Mendoza, Y. and Bello, G. (2014) 'Spatiotemporal dynamics of dissemination of non-pandemic HIV-1 subtype B clades in the caribbean region', *PLoS ONE*, 9(8). doi: 10.1371/journal.pone.0106045.

Chen, J. H., Wong, K., Chan, K. C., To, S. W. and Chen, Z. (2011) 'Phylodynamics of HIV-1 Subtype B among the Men- Having-Sex-with-Men (MSM) Population in Hong Kong', 6(9). doi: 10.1371/journal.pone.0025286.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003) 'Multiple sequence alignment with the Clustal series of programs', *Nucleic Acids Research*, 31(13), pp. 3497–3500. doi: 10.1093/nar/gkg500.

Civetta, A., Ostapchuk, D. C. M. and Nwali, B. (2016) 'Genome hotspots for nucleotide substitutions and the evolution of influenza A (H1N1) human strains.', *Genome biology and evolution*, 8(4), pp. 986–993. doi: 10.1093/gbe/evw061.

Delva, W. and Karim, Q. A. (2014) 'The HIV Epidemic in Southern Africa – Is an AIDS-Free Generation Possible ?', pp. 99–108. doi: 10.1007/s11904-014-0205-0.

Engelbrecht, S., Laten, J. D., Smith, T. L. and van Rensburg, E. J. (1995) 'Identification of env subtypes in fourteen HIV type 1 isolates from south Africa.', *AIDS research and human retroviruses*, 11(10), pp. 1269–1271.

Faria, N. R., Rambaut, A., Suchard, M. A., Baele, G., Bedford, T., Ward, M. J., Tatem, A. J., Sousa, J. D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O. G. and Lemey, P. (2014) 'Supplementary Materials for The early spread and epidemic ignition of HIV-1 in human populations', *Science*, 346(6205), pp. 56–61. doi: 10.1126/science.1256739.

Felsenstein, J. (1985) 'Confidence Limits on Phylogenies : An Approach Using the Bootstrap Author (s): Joseph Felsenstein Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/2408678', *Evolution*, 39(4), pp. 783–791.

Fraser-Hurt, N., Zuma, K., Njuho, P., Chikwava, F., Slaymaker, E., Hosegood, V. and Gorgens, M.

(2011) ‘The HIV epidemic in South Africa : What do we know and how has it changed?’, p. 74. Available at: http://www.hsrc.ac.za/en/research-outputs/ktree-doc/9611.

Freed, E. O. (2015) ‘HIV-1 assembly, release and maturation’, *Nat Rev Microbiol*. Nature Publishing Group, 13(8), pp. 484–496. doi: 10.1038/nrmicro3490.

Friedman-Kien, A. E. (1981) ‘Disseminated Kaposi â€™ s sarcoma syndrome in young homosexual men’, *Journal of the American Academy of Dermatology*, 5(4), pp. 648–671. doi: 10.1016/S0190-9622(81)80010-2.

Gall, A., Ferns, B., Morris, C., Watson, S., Cotten, M., Robinson, M., Berry, N., Pillay, D. and Kellam, P. (2012) ‘Universal amplification, next-generation sequencing, and assembly of HIV-1 genomes’, *Journal of Clinical Microbiology*, 50(12), pp. 3838–3844. doi: 10.1128/JCM.01516-12.

Gallo, R. C. (1984) ‘Maryland 20895’, *Science*, 224(38), pp. 500–503.

Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M. and Hahn, B. H. (1999) ‘Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes’, *Nature*, 397(6718), pp. 436–441. doi: 10.1038/17130.

Gao, F., Robertson, D. L., Morrison, S. G., Hui, H., Craig, S., Decker, J., Fultz, P. N., Girard, M., Shaw, G. M., Hahn, B. H. and Sharp, P. M. (1996) ‘The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin.’, *Journal of virology*, 70(10), pp. 7013–29.

Gilbert, M. T. P., Rambaut, A., Wlasiuk, G., Spira, T. J., Pitchenik, A. E. and Worobey, M. (2007) ‘The emergence of HIV/AIDS in the Americas and beyond.’, *Proceedings of the National Academy of Sciences of the United States of America*, 104(47), pp. 18566–18570. doi: 10.1073/pnas.0705329104.

Girard, M. P., Osmanov, S., Assossou, O. M. and Kieny, M.-P. (2011) ‘Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: A review.’, *Vaccine*. Elsevier Ltd, 29(37), pp. 6191–6218. doi: 10.1016/j.vaccine.2011.06.085.

Gordon, M., De Oliveira, T., Bishop, K., Coovadia, H. M., Madurai, L., Engelbrecht, S., Janse van Rensburg, E., Mosam, A., Smith, A. and Cassol, S. (2003) ‘Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies.’, *Journal of virology*, 77(4), pp. 2587–99. doi: 10.1128/JVI.77.4.2587.

Graves, A. H., Hahn, B. H., Gao, F., Yue, L., Hill, S. C., David, L., Saag, M. S., Shaw, G. M. and Paul, M. (1994) ‘H ’, 10(5), pp. 625–627.

Graves, M. C., Lim, J. J., Heimer, E. P. and Kramer, R. A. (1988) 'An 11-kDa form of human immunodeficiency virus protease expressed in Escherichia coli is sufficient for enzymatic activity', *Proceedings of the National Academy of Sciences*, 85(8), pp. 2449–2453. doi: 10.1073/pnas.85.8.2449.

Grossmann, S., Nowak, P. and Neogi, U. (2015) 'Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management', *Journal of the International AIDS Society*, 18(1), pp. 1–8. doi: 10.7448/IAS.18.1.20035.

Guimara, M. L., Bello, G., Eyer-silva, W. A., Chequer-fernandez, S. L., Teixeira, S. L. M. and Morgado, M. G. (2007) 'Demographic history of HIV-1 subtypes B and F in Brazil', 7, pp. 263–270. doi: 10.1016/j.meegid.2006.11.002.

Hahn, B. H., Shaw, G. M., Cock, K. M. De and Sharp, P. M. (2000) 'AIDS as a Zoonosis : Scientific and Public Health Implications', 287(January).

Harmelen, J. H. V. A. N., Ryst, E. V. A. N. D. E. R., Loubser, A. S., York, D. and Madurai, S. (1999) 'A Predominantly HIV Type 1 Subtype C-Restricted Epidemic in South African Urban Populations', 15(4), pp. 395–398.

Harmelen, J. V. A. N., Williamson, C., Kim, B., Morris, L., Carr, J., Karim, S. S. A. and Cutchan, F. M. C. (2001) 'Characterization of Full-Length HIV Type 1 Subtype C Sequences from South Africa', *AIDS research and human retroviruses*, 17(16), pp. 1527–1531.

Van Harmelen, J., Wood, R., Lambrick, M., Rybicki, E. P., Williamson, A.-L. and Williamson, C. (1997) 'An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa', *Aids*, 11(1), pp. 81–87. doi: 10.1097/00002030-199701000-00012.

Hawke, K. G., Waddell, R. G., Gordon, D. L., Ratcliff, R. M., Ward, P. R. and Kaldor, J. M. (2013) 'HIV Non-B Subtype Distribution : Emerging Trends and Risk Factors for Imported and Local Infections', 29(2). doi: 10.1089/aid.2012.0082.

Hemelaar, J., Gouws, E., Ghys, P. D. and Osmanov, S. (2011) 'Global trends in molecular epidemiology of HIV-1 during 2000– 2007', *AIDS (London, England)*, 25(5), pp. 679–689. doi: 10.1097/QAD.0b013e328342ff93.Global.

Hemelaar, J., Gouws, E., Ghys, P. D., Osmanov, S. and WHO-UNAIDS Network for HIV Isolation and Characterisation (2011) 'Global trends in molecular epidemiology of HIV-1 during 2000-2007.', *AIDS (London, England)*, 25(5), pp. 679–89. doi: 10.1097/QAD.0b013e328342ff93.

Holmes, M., Zhang, F. and Bieniasz, P. D. (2015) 'Single-Cell and Single-Cycle Analysis of HIV-1

Replication', *PLOS Pathogens*, 11(6), p. e1004961. doi: 10.1371/journal.ppat.1004961.

Hunt, G. M., Papathanasopoulos, M. A., Gray, G. E. and Tiemessen, C. T. (2003) 'Characterisation of near-full length genome sequences of three South African human immunodeficiency virus type 1 subtype C isolates', *Virus Genes*, 26(1), pp. 49–56. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12680693.

J., H. (2012) 'The origin and diversity of the HIV-1 pandemic.', *Trends in Molecular Medicine*. Elsevier Ltd, 18(3), pp. 182–192. doi: 10.1016/j.molmed.2011.12.001.

Jacobs, G. B., Loxton, A. G., Laten, A. and Engelbrecht, S. (2007) 'Complete genome sequencing of a non-syncytium-inducing HIV type 1 subtype D strain from Cape Town, South Africa.', *AIDS research and human retroviruses*, 23(12), pp. 1575–1578. doi: 10.1089/aid.2007.0167.

Jacobs, G. B., Nistal, M., Laten, A., van Rensburg, E. J., Rethwilm, A., Preiser, W., Bodem, J. and Engelbrecht, S. (2008) 'Molecular analysis of HIV type 1 vif sequences from Cape Town, South Africa.', *AIDS research and human retroviruses*, 24(7), pp. 991–4. doi: 10.1089/aid.2008.0077.

Jacobs, G. B., Wilkinson, E., Isaacs, S., Spies, G., Oliveira, T. De, Seedat, S. and Engelbrecht, S. (2014) 'HIV-1 Subtypes B and C Unique Recombinant Forms (URFs) and Transmitted Drug Resistance Identified in the Western Cape Province , South Africa', 9(3). doi: 10.1371/journal.pone.0090845.

Jaskólski, M., Tomasselli, a G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Schneider, J., Kent, S. B. and Wlodawer, A. (1991) 'Structure at 2.5-A resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor.', *Biochemistry*, 30(6), pp. 1600–1609. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1993177.

Junqueira, D. M. and de Matos Almeida, S. E. (2016) 'HIV-1 subtype B: Traces of a pandemic', *Virology*. Elsevier, 495, pp. 173–184. doi: 10.1016/j.virol.2016.05.003.

Junqueira, D. M., de Medeiros, R. M., Matte, M. C. C., Araújo, L. A. L., Chies, J. A. B., Ashton-Prolla, P. and de Matos Almeida, S. E. (2011) 'Reviewing the history of HIV-1: Spread of subtype B in the Americas', *PLoS ONE*, 6(11). doi: 10.1371/journal.pone.0027489.

Kato, S., Saito, Y., Tanaka, R., Hiraishi, Y., Kitamura, N., Matsumoto, T., Hanabusa, H., Kamakura, M., Ikeda, Y. and Negishi, M. (2003) 'Differential prevalence of HIV type 1 subtype B and CRF01_AE among different sexual transmission groups in Tokyo, Japan, as revealed by subtype-specific PCR', *Aids Research and Human Retroviruses*, 19(11), pp. 1057–1063. doi: Doi 10.1089/088922203322588431.

Keele, B. F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M. L., Bibollet-Ruche, F., Chen, Y., Wain, L. V, Liegeois, F., Loul, S., Ngole, E. M., Bienvenue, Y., Delaporte, E., Brookfield, J. F. Y., Sharp, P. M., Shaw, G. M., Peeters, M. and Hahn, B. H. (2006) 'Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.', *Science (New York, N.Y.)*, 313(5786), pp. 523–6. doi: 10.1126/science.1126531.

Kondo, M., Lemey, P., Sano, T., Itoda, I., Yoshimura, Y., Sagara, H., Tachikawa, N., Yamanaka, K., Iwamuro, S., Matano, T., Imai, M., Kato, S. and Takebe, Y. (2013) 'Emergence in Japan of an HIV-1 Variant Associated with Transmission among Men Who Have Sex with Men (MSM) in China ':, 87(10), pp. 5351–5361. doi: 10.1128/JVI.02370-12.

Kopietz, F., Jaguva Vasudevan, a. a., Kramer, M., Muckenfuss, H., Sanzenbacher, R., Cichutek, K., Flory, E. and Munk, C. (2012) 'Interaction of human immunodeficiency virus type 1 Vif with APOBEC3G is not dependent on serine/threonine phosphorylation status', *Journal of General Virology*, 93(Pt_11), pp. 2425–2430. doi: 10.1099/vir.0.043273-0.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B. H., Wolinsky, S. and Bhattacharya, T. (2000) 'Timing the ancestor of the HIV-1 pandemic strains.', *Science (New York, N.Y.)*, 288(5472), pp. 1789–1796. doi: 10.1126/science.288.5472.1789.

Kuiken, C., Thakallapalli, R., Eskild, A. and De Ronde, A. (2000) 'Genetic analysis reveals epidemiologic patterns in the spread of human immunodeficiency virus', *American Journal of Epidemiology*, 152(9), pp. 814–822. doi: 10.1093/aje/152.9.814.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007) 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23(21), pp. 2947–2948. doi: 10.1093/bioinformatics/btm404.

Lau, K. A. and Wong, J. J. L. (2013) 'Current trends of HIV recombination worldwide', *Infectious Disease Reports*, 5(SUPPL.1), pp. 15–20. doi: 10.4081/idr.2013.s1.e4.

Laverdiere, M., Tremblay, J. and Lavallee, R. (1983) 'AIDS in Haitian immigrants and in a Caucasian woman closely associated with Haitians', *Canadian Medical Association Journal*, 129(11), pp. 1209–1212.

Lemey, P., Pybus, O. G., Rambaut, A., Drummond, A. J., Robertson, D. L., Roques, P., Worobey, M. and Vandamme, A. M. (2004) 'The molecular population genetics of HIV-1 group O', *Genetics*, 167(3), pp. 1059–1068. doi: 10.1534/genetics.104.026666.

Liebert, M. A., Casado, C., Urtasun, I., Saragosti, S., Chaix, M., Rossi, A. D. E., Cattelan, A. M. and

Dietrich, U. (2000) 'Different Distribution of HIV Type 1 Genetic Variants in European Patients with Distinct Risk Practices', 16(3), pp. 299–304.

Loxton, A. G., Treurnicht, F., Laten, A., van Rensburg, E. J. and Engelbrecht, S. (2005) 'Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986.', *AIDS research and human retroviruses*, 21(5), pp. 410–413. doi: 10.1089/aid.2005.21.410.

Lukashov, V. V, Kuiken, C. L., Vlahov, D., Coutinho, R. A. and Goudsmit, J. (1996) 'Evidence for HIV type 1 strains of U.S. intravenous drug users as founders of AIDS epidemic among intravenous drug users in Northern Europe', *AIDS Research and Human Retroviruses*, 12(12), pp. 1179–1183. doi: 10.1089/aid.1996.12.1179.

Magiorkinis, G., Angelis, K., Mamais, I., Katzourakis, A., Hatzakis, A., Albert, J., Lawyer, G., Hamouda, O., Struck, D., Vercauteren, J., Wensing, A., Alexiev, I., Åsjö, B., Balotta, C., Gomes, P., Camacho, R. J., Coughlan, S., Griskevicius, A., Grossman, Z., Horban, A., Kostrikis, L. G., Lepej, S. J., Liitsola, K., Linka, M., Nielsen, C., Otelea, D., Paredes, R., Poljak, M., Puchhammer-Stöckl, E., Schmit, J. C., Sönnerborg, A., Staneková, D., Stanojevic, M., Boucher, C. A. B., SPREAD program, Nikolopoulos, G., Vasylyeva, T., Friedman, S. R., van de Vijver, D., Angarano, G., Chaix, M.-L., de Luca, A., Korn, K., Loveday, C., Soriano, V., Yerly, S., Zazzi, M., Vandamme, A.-M. and Paraskevis, D. (2016) 'TEMPORARY REMOVAL: The global spread of HIV-1 subtype B epidemic.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. Elsevier B.V. doi: 10.1016/j.meegid.2016.05.041.

Malebranche, R., Guérin, J. ., Laroche, A. ., Elie, R., Spira, T., Drotman, P., Arnoux, E., Pierre, G. ., Pean-Guichard, C., Morisset, P. ., Mandeville, R., Seemayer, T. and Dupuy, J.-M. (1983) 'Acquired Immunodeficiency Syndrome with severe gastrointestinal manifestaions in Haiti', *The Lancet*, 322(8355), pp. 873–878. doi: 10.1016/S0140-6736(83)90868-1.

Mascarenhas, A. P. and Musier-Forsyth, K. (2009) 'The capsid protein of human immunodeficiency virus: Interactions of HIV-1 capsid with host protein factors', *FEBS Journal*, 276(21), pp. 6118–6127. doi: 10.1111/j.1742-4658.2009.07315.x.

Mayosi, B. M., Ch, B., Phil, D., Benatar, S. R., Ch, B. and Med, D. S. (2014) 'spe ci a l r e p or t Health and Health Care in South Africa — 20 Years after Mandela'.

zur Megede, J., Engelbrecht, S., de Oliveira, T., Cassol, S., Scriba, T. J., van Rensburg, E. J. and Barnett, S. W. (2002) 'Novel evolutionary analyses of full-length HIV type 1 subtype C molecular clones from Cape Town, South Africa.', *AIDS research and human retroviruses*, 18(17), pp. 1327–

32. doi: 10.1089/088922202320886370.

Middelkoop, K., Rademeyer, C., Brown, B. B., Cashmore, T. J., Marais, J. C., Scheibe, A. P., Bandawe, G. P., Myer, L., Fuchs, J. D., Williamson, C. and Bekker, L. G. (2014) 'Epidemiology of HIV-1 subtypes among men who have sex with men in Cape Town, South Africa', *J Acquir Immune Defic Syndr*, 65(4), pp. 473–480. doi: 10.1097/qai.0000000000000067.

Nabatov, A. A., Kravchenko, O. N., Lyulchuk, M. G., Shcherbinskaya, A. M. and Lukashov, V. V (2002) 'Simultaneous introduction of HIV type 1 subtype A and B viruses into injecting drug users in southern Ukraine at the beginning of the epidemic in the former Soviet Union', *AIDS Research and Human Retroviruses*, 18(12), pp. 891–895. doi: 10.1089/08892220260190380.

Neogi, U., Bontell, I., Shet, A., de Costa, A., Gupta, S., Diwan, V., Laishram, R. S., Wanchu, A., Ranga, U., Banerjea, A. C. and Sönnerborg, A. (2012) 'Molecular epidemiology of HIV-1 subtypes in India: Origin and evolutionary history of the predominant subtype C', *PLoS ONE*, 7(6). doi: 10.1371/journal.pone.0039819.

Ng, K. T., Ong, L. Y., Lim, S. H., Takebe, Y., Kamarulzaman, A. and Tee, K. K. (2013) 'Evolutionary History of HIV-1 Subtype B and CRF01 _ AE Transmission Clusters among Men Who Have Sex with Men (MSM) in Kuala Lumpur , Malaysia', 8(6). doi: 10.1371/journal.pone.0067286.

Novitsky, V. A., Montano, M. A., Mclane, M. F., Renjifo, B., Vannberg, F., Foley, B. T., Thumbi, P., Rahman, M., Makhema, M. J., Essex, M., Novitsky, V. A., Montano, M. A., Lane, M. F. M. C., Renjifo, B., Vannberg, F., Foley, B. T., U, T. P. N. and Rahman, M. (1999) 'Molecular Cloning and Phylogenetic Analysis of Human Immunodeficiency Virus Type 1 Subtype C : a Set of 23 Full-Length Clones from Botswana Molecular Cloning and Phylogenetic Analysis of Human Immunodeficiency Virus Type 1 Subtype C : a Set of 23 Full-Len', 73(5), pp. 4427–4432.

Pagán, I. and Holguín, Á. (2013) 'Reconstructing the Timing and Dispersion Routes of HIV-1 Subtype B Epidemics in The Caribbean and Central America: A Phylogenetic Story', *PLoS ONE*, 8(7). doi: 10.1371/journal.pone.0069218.

Paraskevis, D., Pybus, O., Magiorkinis, G., Hatzakis, A., Wensing, A. M., van de Vijver, D. A., Albert, J., Angarano, G., Asjö, B., Balotta, C., Boeri, E., Camacho, R., Chaix, M.-L., Coughlan, S., Costagliola, D., De Luca, A., de Mendoza, C., Derdelinckx, I., Grossman, Z., Hamouda, O., Hoepelman, I., Horban, A., Korn, K., Kücherer, C., Leitner, T., Loveday, C., Macrae, E., Maljkovic-Berry, I., Meyer, L., Nielsen, C., Op de Coul, E. L., Ormaasen, V., Perrin, L., Puchhammer-Stöckl, E., Ruiz, L., Salminen, M. O., Schmit, J.-C., Schuurman, R., Soriano, V., Stanczak, J., Stanojevic, M., Struck, D., Van Laethem, K., Violin, M., Yerly, S., Zazzi, M., Boucher, C. A. and Vandamme,

A.-M. (2009) 'Tracing the HIV-1 subtype B mobility in Europe: a phylogeographic approach.', *Retrovirology*, 6, p. 49. doi: 10.1186/1742-4690-6-49.

Patel, P., Borkowf, C. B., Brooks, J. T., Lasry, A., Lansky, A. and Mermin, J. (2014) 'Estimating per-act HIV transmission risk : a systematic review', (April). doi: 10.1097/QAD.0000000000000298.

Perrin, L., Kaiser, L. and Yerly, S. (2003) 'Travel and the spread of HIV-1 genetic variants', *Lancet Infectious Diseases*, 3(1), pp. 22–27. doi: 10.1016/S1473-3099(03)00484-5.

Piot, P., Taelman, H., Bila Minlangu, K., Mbendi, N., Ndangi, K., Kalambayi, K., Bridts, C., Quinn, T. C., Feinsod, F. M., Wobin, O., Mazebo, P., Stevens, W., Mitchell, S. and Mccormick, J. B. (1984) 'Acquired Immunodeficiency Syndrome in a Heterosexual Population in Zaire', *The Lancet*, 324(8394), pp. 65–69. doi: 10.1016/S0140-6736(84)90241-1.

Plantier, J.-C., Plantier, J.-C., Leoz, M., Leoz, M., Dickerson, J. E., Dickerson, J. E., De Oliveira, F., De Oliveira, F., Cordonnier, F., Cordonnier, F., Lemée, V., Lemée, V., Damond, F., Damond, F., Robertson, D. L., Robertson, D. L., Simon, F. and Simon, F. (2009) 'A new human immunodeficiency virus derived from gorillas.', *Nature medicine*, 15(8), pp. 871–2. doi: 10.1038/nm.2016.

Pomerantz, R. J. and Horn, D. L. (2003) 'Twenty years of therapy for HIV-1 infection', *Nature medicine*, 9(7), pp. 867–873. doi: 10.1038/nm0703-867.

Puren, A, J. (2002) 'The HIV-1 epidemic in South Africa.', *Oral diseases*, 8 Suppl 2, pp. 27–31. doi: 10.1034/j.1601-0825.2002.00007.x.

Robbins, K. E., Lemey, P., Pybus, O. G., Jaffe, H. W., Youngpairoj, A. S., Brown, T. M., Salemi, M., Vandamme, A.-M. and Kalish, M. L. (2003) 'U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains.', *Journal of virology*, 77(11), pp. 6359–66. doi: 10.1128/JVI.77.11.6359.

Robertson, D. L., Anderson, J. P., Bradac, J. a, Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000a) 'HIV-1 nomenclature proposal.', *Science (New York, N.Y.)*, 288(5463), pp. 492–505. doi: 10.1126/science.288.5463.55d.

Robertson, D. L., Anderson, J. P., Bradac, J. a, Carr, J. K., Foley, B., Funkhouser, R. K., Gao, F., Hahn, B. H., Kalish, M. L., Kuiken, C., Learn, G. H., Leitner, T., McCutchan, F., Osmanov, S., Peeters, M., Pieniazek, D., Salminen, M., Sharp, P. M., Wolinsky, S. and Korber, B. (2000b) 'HIV-1 nomenclature proposal.', *Science (New York, N.Y.)*, 288(5463), pp. 55–56. doi: 10.1126/science.288.5463.55d.

Rousseau, C. M., Birditt, B. A., McKay, A. R., Stoddard, J. N., Lee, T. C., McLaughlin, S., Moore, S. W., Shindo, N., Learn, G. H., Korber, B. T., Brander, C., Goulder, P. J. R., Kiepiela, P., Walker, B. D. and Mullins, J. I. (2006) 'Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes', *Journal of Virological Methods*, 136(1–2), pp. 118–125. doi: 10.1016/j.jviromet.2006.04.009.

Saitou, N. and Nei, M. (1987) 'The neighbour-joining method: a new method for reconstructing phylogenetic trees', *Molecular Biology Evolution*, 4(4), pp. 406–425. doi: citeulike-article-id:93683.

Selik H. W.; Curran, J. W., R. M. . H. (1984) 'Acquired immune deficiency syndrome (AIDS) trends in the United States, 1978-1982', *Am J Med*, 76(3), pp. 493–500.

Sharp, P. M., Bailes, E., Chaudhuri, R. R., Rodenburg, C. M., Santiago, M. O. and Hahn, B. H. (2001) 'The origins of acquired immune deficiency syndrome viruses: where and when?', *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 356(1410), pp. 867–876. doi: 10.1098/rstb.2001.0863.

Sher, R. (1989) 'HIV infection in South Africa, 1982-1988--a review', *South African medical journal = Suid-Afrikaanse tydskrif vir geneeskunde*, 76(7), p. 314—318. Available at: http://europepmc.org/abstract/MED/2799575.

SIEPEL, A. C., HALPERN, A. L., MACKEN, C. and KORBER, B. T. M. (1995) 'A Computer Program Designed to Screen Rapidly for HIV Type 1 Intersubtype Recombinant Sequences', *AIDS Research and Human Retroviruses*, 11(11), pp. 1413–1416. doi: 10.1089/aid.1995.11.1413.

Society, I. B. (2010) 'Minimum Mutation Fits to a Given Tree Author (s): J . A . Hartigan Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529676', 29(1), pp. 53–65.

Solbak, S. M. Ø., Reksten, T. R., Hahn, F., Wray, V., Henklein, P., Henklein, P., Halskau, Ø., Schubert, U. and Fossen, T. (2013) 'HIV-1 p6 - A structured to flexible multifunctional membrane-interacting protein', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier B.V., 1828(2), pp. 816–823. doi: 10.1016/j.bbamem.2012.11.010.

Stanley, B. J., Ehrlich, E. S., Short, L., Yu, Y., Xiao, Z., Yu, X.-F. and Xiong, Y. (2008) 'Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly.', *Journal of virology*, 82(17), pp. 8656–63. doi: 10.1128/JVI.00767-08.

Steel, M. and Penney, D. (2000) 'Parsimony, likelihood, and the role of models in molecular phylogenetics.', *Molecular Biology and Evolution*, 17, pp. 839–850. doi: 10.1093/oxfordjournals.molbev.a026364.

Takahashi, K. and Nei, M. (2000) 'Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used.', *Molecular biology and evolution*, 17(8), pp. 1251–1258. doi: 10.1093/oxfordjournals.molbev.a026408.

Takebe, Y., Naito, Y., Raghwani, J., Fearnhill, E., Sano, T., Kusagawa, S., Mbisa, J. L., Zhang, H., Matano, T., Brown, A. J. L., Pybus, O. G., Dunn, D. and Kondo, M. (2014a) 'Intercontinental dispersal of HIV-1 subtype B associated with transmission among men who have sex with men in Japan.', *Journal of virology*, 88(17), pp. 9864–76. doi: 10.1128/JVI.01354-14.

Takebe, Y., Naito, Y., Raghwani, J., Fearnhill, E., Sano, T., Kusagawa, S., Mbisa, J. L., Zhang, H., Matano, T., Brown, A. J. L., Pybus, O. G., Dunn, D. and Kondo, M. (2014b) 'Intercontinental dispersal of HIV-1 subtype B associated with transmission among men who have sex with men in Japan.', *Journal of virology*, 88(17), pp. 9864–76. doi: 10.1128/JVI.01354-14.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) 'MEGA6: Molecular evolutionary genetics analysis version 6.0', *Molecular Biology and Evolution*, 30(12), pp. 2725–2729. doi: 10.1093/molbev/mst197.

Tebit, D. M. and Arts, E. J. (2011) 'Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease', *The Lancet Infectious Diseases*. Elsevier Ltd, 11(1), pp. 45–56. doi: 10.1016/S1473-3099(10)70186-9.

Tee, K. K., Saw, T. L., Pon, C. K., Kamarulzaman, A. and Ng, K. P. (2005) 'The evolving molecular epidemiology of HIV type 1 among injecting drug users (IDUs) in Malaysia', *AIDS Research and Human Retroviruses*, 21(12), pp. 1046–1050. Available at: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L43069188%5Cn http://dx.doi.org/10.1089/aid.2005.21.1046.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice', *Nucleic Acids Research*, 22(22), pp. 4673–4680. doi: 10.1093/nar/22.22.4673.

Thomson, M. M., Pérez-álvarez, L. and Nájera, R. (2002) 'Reviews Molecular epidemiology of HIV-1 genetic forms therapy', *The Lancet*, 2, pp. 461–471.

Treurnicht, F. K., Smith, T.-L., Engelbrecht, S., Claassen, M., Robson, B. a, Zeier, M. and van Rensburg, E. J. (2002) 'Genotypic and phenotypic analysis of the env gene from South African HIV-1 subtype B and C isolates.', *Journal of medical virology*, 68(2), pp. 141–6. doi: 10.1002/jmv.10199.

Turner, B. G. and Summers, M. F. (1999) 'Structural biology of HIV', *Journal of Molecular Biology*, 285(1), pp. 1–32. doi: 10.1006/jmbi.1998.2354.

Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembi, N., Gürtler, L., Devare, S. and Brennan, C. A. (2011) 'Confirmation of putative HIV-1 group P in Cameroon.', *Journal of virology*, 85(3), pp. 1403–7. doi: 10.1128/JVI.02005-10.

Weiss, R. A. (2006) 'The discovery of endogenous retroviruses', 11, pp. 1–11. doi: 10.1186/1742-4690-3-67.

Wheeler, V. W. and Radcliffe, K. W. (1994) 'HIV infection in the Caribbean', *Int J STD AIDS*, 5(2), pp. 79–89. doi: 10.1177/095646249400500201.

Wilkinson, E. and Engelbrecht, S. (2009) 'Molecular characterization of non-subtype C and recombinant HIV-1 viruses from Cape Town, South Africa', *Infection, Genetics and Evolution*, 9(5), pp. 840–846. doi: 10.1016/j.meegid.2009.05.001.

Wilkinson, E., Holzmayer, V., Jacobs, G. B., de Oliveira, T., Brennan, C. A., Hackett Jr., J., van Rensburg, E. J. and Engelbrecht, S. (2015) 'Sequencing and Phylogenetic Analysis of Near Full-Length HIV-1 Subtypes A, B, G and Unique Recombinant AC and AD Viral Strains Identified in South Africa', *Aids Research and Human Retroviruses*, 31(4), pp. 412–420. doi: 10.1089/aid.2014.0230.

Worobey, M. (2004) 'Origin of AIDS: contaminated polio vaccine theory refuted', *Nature*, 428(April), p. 820. doi: 10.1038/428820a.

Worobey, M., Gemmel, M., Teuwen, D. E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.-J., Kabongo, J.-M. M., Kalengayi, R. M., Van Marck, E., Gilbert, M. T. P. and Wolinsky, S. M. (2008) 'Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960.', *Nature*, 455(7213), pp. 661–4. doi: 10.1038/nature07390.

Worobey, M., Watts, T. D., McKay, R. A., Suchard, M. A., Granade, T., Teuwen, D. E., Koblin, B. A., Heneine, W., Lemey, P. and Jaffe, H. W. (2016) '1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America', *Nature*. Nature Publishing Group, pp. 1–17. doi: 10.1038/nature19827.

Yoshida, R. and Nei, M. (2016) 'Efficiencies of the NJp, Maximum Likelihood, and Bayesian Methods of Phylogenetic Construction for Compositional and Noncompositional Genes', *Molecular Biology and Evolution*, 33(6), pp. 1618–1624. doi: 10.1093/molbev/msw042.

Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M. and Ho, D. D. (1998) 'An African HIV-1 sequence from 1959 and implications for the origin of the epidemic', *Nature*, 391(6667), pp.

594–597. doi: 10.1038/35400.

Zuma, K., Shisana, O., Rehle, T. M., Simbayi, L. C., Jooste, S., Zungu, N., Labadarios, D., Onoya, D., Evans, M., Moyo, S. and Abdullah, F. (2016) 'New insights into HIV epidemic in South Africa: key findings from the National HIV Prevalence, Incidence and Behaviour Survey, 2012', *African Journal of AIDS Research*, 15(1), pp. 67–75. doi: 10.2989/16085906.2016.1153491.

Appendix One

Ethics approval

UNIVERSITEIT•STELLENBOSCH•UNIVERSITY
jou kennisvennoot • your knowledge partner

Approval Notice
Response to Modifications- (New Application)

13-Oct-2015
Jacobs, Graeme GB

Ethics Reference #: N15/08/071
Title: Tracking the molecular epidemiology and resistance patterns of HIV-1 in South Africa.

Dear Dr. Graeme Jacobs,

The **Response to Modifications** - ***(New Application)*** received on **28-Sep-2015**, was reviewed by members of **Health Research Ethics Committee 1** via Expedited review procedures on **13-Oct-2015** and was approved.
Please note the following information about your approved research protocol:

Protocol Approval Period: **13-Oct-2015 -12-Oct-2016**

Please remember to use your **protocol number** (**N15/08/071**) on any documents or correspondence with the HREC concerning your research protocol.

Please note that the HREC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

After Ethical Review:
Please note a template of the progress report is obtainable on www.sun.ac.za/rds and should be submitted to the Committee before the year has expired. The Committee will then consider the continuation of the project for a further year (if necessary). Annually a number of projects may be selected randomly for an external audit.
Translation of the consent document to the language applicable to the study participants should be submitted.

Federal Wide Assurance Number: 00001372
Institutional Review Board (IRB) Number: IRB0005239

The Health Research Ethics Committee complies with the SA National Health Act No.61 2003 as it pertains to health research and the United States Code of Federal Regulations Title 45 Part 46. This committee abides by the ethical norms and principles for research, established by the Declaration of Helsinki, the South African Medical Research Council Guidelines as well as the Guidelines for Ethical Research: Principles Structures and Processes 2004 (Department of Health).

Provincial and City of Cape Town Approval

Please note that for research at a primary or secondary healthcare facility permission must still be obtained from the relevant authorities (Western Cape Department of Health and/or City Health) to conduct the research as stated in the protocol. Contact persons are Ms Claudette Abrahams at Western Cape Department of Health (healthres@pgwc.gov.za Tel: +27 21 483 9907) and Dr Helene Visser at City Health (Helene.Visser@capetown.gov.za Tel: +27 21 400 3981). Research that will be conducted at any tertiary academic institution requires approval from the relevant hospital manager. Ethics approval is required BEFORE approval can be obtained from these health authorities.

We wish you the best as you conduct your research.
For standard HREC forms and documents please visit: www.sun.ac.za/rds

If you have any questions or need further assistance, please contact the HREC office at 0219399657.

Included Documents:
CV E Obasa
Protocol
Payment instruction form
Protocol Synopsis

Appendix Two

Sequencing primers used to sequence the HIV-1 genome

Sequencing primers	Primers (5'-3')	Bases	Tm (50mM)	Orientation	Reference
gag-pol					
LPgag 6F.1	CAGCCCAGAAGTAATACCCATGTT	24	58.5	Forward	Dr Hackett JR*
R2051	TATRTTGACAGGTGTAGGT	19	49.5	Reverse	Dr Hackett JR
R2457	CTAAGGGRACTGAAAAATATGC	22	51.7	Reverse	Dr Hackett JR
F2840	TGGACTGTCAATGATATACAGA	22	52.1	Reverse	Dr Hackett JR
poli7	AACAAGTAGATAAATTAGTCAGT	23	48.7	Forward	Dr Hackett JR
ABB20-10R	GGCTAGGTGAATTGCATGTA	20	53.6	Forward	Dr Hackett JR
ABB20-11R	TATGTCCATTGGTCTTGCCC	20	55.9	Reverse	Dr Hackett JR
ABB20-3F	ATCAGTACAATGTGCTTCCA	20	52.9	Forward	Dr Hackett JR
Mgag 17R	GCTAKRTGYCCTTCYTTGCCACA	23	61.0	Reverse	Dr Hackett JR
poli2	TAAARACARYAGTACWAATGGCA	23	52.9	Forward	Dr Hackett JR
M/O p24-6R	TGTGWAGCTTGYTCRGCTC	19	56.1	Reverse	Dr Hackett JR
ppr15	CCTTCTAAATGTGTACAATC	20	46.9	Forward	Dr Hackett JR
cm237F2030	GGAAACCAAAAATGATAGGGGG	22	55.2	Forward	Dr Hackett JR
cm237R2500	GTACTGATATCTAATCCCTGG	21	50.0	Reverse	Dr Hackett JR
cm237R3800	CTGACTAATTTATCTACTTG	20	43.5	Reverse	Dr Hackett JR
cm237R2020	GGTGGGGCTGTTGGCTCTGG	20	64.0	Reverse	Dr Hackett JR
ppf13b	AAGATGGCCAGTAAAAGTAATACACACAGACAA	33	61.9	Reverse	Dr Hackett JR
RTUG F3	GAAGCAGAATTAGAAYTGGCAGA	23	55.7	Forward	Dr Hackett JR
rtin seq F2	CAACCAGAYARRAGTGAATCAGA	23	54.9	Forward	Dr Hackett JR
cm237F3000	TGTGRGTCTGTTACTATRTTTACTTC	26	54.1	Forward	Dr Hackett JR
ABB20-2F	TAGGTACAGTGTTAGTAGGA	20	49.7	Forward	Dr Hackett JR
ABB20-6F	GGCATAGATAAAGCCCAAGAA	21	53.8	Forward	Dr Hackett JR

pol3D	CAGTACTGGATGTGGG	16	48.5	Forward	Dept Medical Virology+
LTR-*gag*					
MSF12	AAATCTCTAGCAGTGGCGCCCCGAACAG	28	69.0	Forward	Salminen et al, 1995
G00	GACTAGCGGAGGCTAGAAG	19	53.1	Forward	Sanders-Buell et al, 1995
G10	CAGTATTAAGCGGGGGAGAATT	22	52.8	Forward	Sanders-Buell et al, 1995
G30	CAGTAGCAACCCTCTATTGTGT	22	52.8	Forward	Sanders-Buell et al, 1995
G60	CAGCCAAAATTACCCTATAGTGCAG	25	55.9	Forward	Sanders-Buell et al, 1995
LPgag 6F.2	CAGCCCAGAAGTAATACCCATGTT	24	58.5	Forward	Dr Hackett JR
ABB55-3-263F (re-gag)	GAGAGCGGCGACTGGTGAG	19	60.7	Forward	Dr Hackett JR
ABB55-3-790R	AGGCCTTTTCTTCTACTACTTTTA	24	53.3	Reverse	Dr Hackett JR
M/O p24-6R.1	TGTGWAGCTTGYTCRGCTC	19	56.1	Reverse	Dr Hackett JR
M/O p24-2F	AGRACYTTRAAYGCATGGGT	20	55.8	Forward	Dr Hackett JR

Sequencing Primers	Primers (5`-3`)	Bases	Tm (50mM)	Orientation	Reference
***pol-env* (IDR)**					
pol4277F.1	GCAGTACAGATGGCAGTATTCAT	23	56.4	Forward	Dr Hackett JR*
pol4565F.1	TATGGAAAACAGATGGCAGGTGAT	24	58.3	Forward	Dr Hackett JR
LP7725R.1	GTCCAATGCCAATAAGTCTTGTTC	24	56.1	Reverse	Dr Hackett JR
20env5'2018R	TCCCCWTCACTCTCRTTGCCACTA	24	62.3	Reverse	Dr Hackett JR
20env5'1644R	CCGGGTGRTTCCAGGGCTCTA	21	63.0	Reverse	Dr Hackett JR
20env5'2287F	CTTTGAGCCCATTCCCATACATTA	24	56.7	Forward	Dr Hackett JR
20env-1301F	ACTATGGGGTACCGGTGTGGAGA	23	62.8	Forward	Dr Hackett JR
55env222R	AATCGCAAAACCAGCTGGAGCAC	23	62.3	Reverse	Dr Hackett JR
55env173R	TTGCCTTGGTGGGTGCTATTCCTA	24	62.1	Reverse	Dr Hackett JR
55env573R	ATTTCCCTTTCCCATTGTATCCA	23	56.0	Reverse	Dr Hackett JR
55env-5418F	ATCATCCGGGGAGTCAGCCTAAGA	24	62.8	Forward	Dr Hackett JR
env27F	AARCCTCCTACTATCATTATRA	22	49.4	Forward	Dr Hackett JR
Menv19R	CTGGYATAGTGCARCARCA	19	54.9	Reverse	Dr Hackett JR
E0	TAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTA	33	66.8	Forward	Sanders-Buell et al, 1995
E00	TAGAAAGAGCAGAAGACAGTGGCAATGA	28	58.3	Forward	Sanders-Buell et al, 1995
E30	GTGTACCCACAGACCCCAGCCCACAAG	27	65.7	Forward	Sanders-Buell et al, 1995
E70	GGGATCAAAGCCTAAAGCCATGTGTAA	27	58.1	Forward	Sanders-Buell et al, 1995
E80	CCAATTCCCATACATTATTGTG	22	49.1	Forward	Sanders-Buell et al, 1995
E180	GTCTGGTATAGTGCAACAGCA	21	52.2	Forward	Sanders-Buell et al, 1995
E55	GCCCCAGACTGTGAGTTGCAACAGATG	27	62.6	Reverse	Sanders-Buell et al, 1995
E65	AGTGCTTCCTGCTGCTCC	18	52.4	Reverse	Sanders-Buell et al, 1995
E145	CAGCAGTTGAGTTGATACTACTGG	24	55.5	Reverse	Sanders-Buell et al, 1995
***env*-LTR**					
ABB55-4-370F	AACCCGACAGGCCCGAAAGAA	21	62.7	Forward	Dr Hackett JR
ABB55-4-1186F	CCAGGGCCAGGGGTTAGAT	19	59.7	Forward	Dr Hackett JR
20LTR-3825F	TGGGTGGCAAGTGGTCAAAAAGTA	24	60.9	Forward	Dr Hackett JR
69env-4109F	ACATGGGTGGCAAGTGGTCAAAA	23	61.5	Forward	Dr Hackett JR
55-4-1912R	CTTTCGGGCCTGTCGGGTTCC	21	63.7	Reverse	Dr Hackett JR
20env 4038R	GTACCTGCGGCCTGACTGGA	20	62.4	Reverse	Dr Hackett JR
E220	TATCAAAATGGCTGTGGTATATAA	24	48.7	Forward	Sanders-Buell et al, 1995
E250	GGAGGCTTGATAGGTTTAAGAATA	24	52.1	Forward	Sanders-Buell et al, 1995
E01	TCCAGTCCCCCCTTTTCTTTTAAAAA	26	54.7	Reverse	Sanders-Buell et al, 1995

E270	GTGGAACTTCTGGGACGCAG	20	55.7	Forward	Sanders-Buell et al, 1995
NefF	CCTAGAAGAATAACACAGGGCTT	23	55.3	Forward	Dept Medical Virology+
NefSF	TGGATGCTGCTTCAAGCTACT	21	57.6	Forward	Dept Medical Virology
In House Design Gene specific primers					
1460R	ACTATACATTCTTACTATTTTATTTAACCCCA	32	52.3	Reverse	Dept Medical Virology
4730F	GGCTGAACATCTTAAGACAGCAGTACAAATGGC	35	56.7	Forward	Dept Medical Virology
5520R	CCTAGTGGGATGTGTACTTCTGAAC	25	50.5	Reverse	Dept Medical Virology
4750F	GACAGCAGTACAAATGGCAGTATTCATCCA	32	60.5	Forward	Dept Medical Virology
5200R	CTTCTGAACTTATTCTTGGATTAGTAC	29	65.0	Reverse	Dept Medical Virology
5400F	GAAAGGCCATATTAGGACATAGAG	24	55.5	Forward	Dept Medical Virology
6060R	GTGGCATTACATGTACTACTTAC	23	50.0	Reverse	Dept Medical Virology
6950R	CTGTGCTGACATTTGTACATGATCC	25	53.2	Reverse	Dept Medical Virology
6500F	GAAGATTTTAAYATGTGGAAA	21	55.0	Forward	Dept Medical Virology
7390R	GTTTAATAGTACTTGGAATGATA	23	55.0	Reverse	Dept Medical Virology
4539F	GGAGAAGCCATGCATGGACAAGTAG	26	50.2	Forward	Dept Medical Virology
5200R	GTACTTCTGAACTTATTCTTGGAT	23	62.0	Reverse	Dept Medical Virology
4830F	GTAGACATAATAGCAACAGACATAC	25	54.5	Forward	Dept Medical Virology
5160F	AGCTAAGGGATGGTCTTATAGACATCAC	28	55.0	Forward	Dept Medical Virology
5350F	GCAGACCAACTAATTCATCTGTATTAC	26	65.0	Forward	Dept Medical Virology
6310R	GACTGTGACCCACAAATTTTCTGCAGCAC	28	58.5	Reverse	Dept Medical Virology
6880R	CCAGCAGGGGCACAATAATGTATGG	26	56.0	Reverse	Dept Medical Virology
7090F	CGAATCTGTAGTAATTAATTGTACAAGA	27	55.0	Forward	Dept Medical Virology
7790R	GCTCCCAAGAACCCAAGGAACAAAGCTCC	29	58.0	Reverse	Dept Medical Virology
8240F	GCATAACAAACTGGCTGTGG	19	59.0	Forward	Dept Medical Virology
8980R	TCCTCTTGTGCTTCTAGCCAGGCAC	25	62.0	Reverse	Dept Medical Virology
3100F	CATGGATGATTTGTATGTAGGATC	24	56.0	Forward	Dept Medical Virology

3920R	TTAGTAACATATCCTGCTTTTCCTA	25	55.0	Reverse	Dept Medical Virology
4340F	GTAGCCAGCTGTGATAAATGTCAGT	25	52.0	Forward	Dept Medical Virology
4810F	GTGCAGGGGAAAGAATAGTAGACATA	27	55.0	Forward	Dept Medical Virology
5310F	AGAGATATAGCACACAAGTAGACCCT	26	54.0	Forward	Dept Medical Virology
5740R	AATTCTTATTATGGCTTCCACTCCTG	24	60.0	Reverse	Dept Medical Virology
5920F	GGTGTTGCTTTCATTGCCAAGTTTG	24	54.0	Forward	Dept Medical Virology
6540R	CCTCATGCATCTGTTCTACCATGT	23	55.0	Reverse	Dept Medical Virology
8650R	AGCACTATTCTTTAGTTCCTGACTCCA	26	62.0	Reverse	Dept Medical Virology
8270F	ATTCATAATGATAGTAGGAGGCTTG	25	53.5	Forward	Dept Medical Virology
8050F	GTGCCTTGGAATGCTAGTTGGAGTA	25	55.7	Forward	Dept Medical Virology
5340F	GACCTAGCAGACCAACTAATTCACCTGC	27	60.0	Forward	Dept Medical Virology

Primer sequences kindly provided by Dr Hackett (Abbott Laboratories, Illinois, USA) before publication. Primers obtained from the Department of Medical Virology, Tygerberg campus, University of Stellenbosch.